

IVAN DE OLIVEIRA

**FUNÇÃO DISCRIMINANTE QUADRÁTICA APLICADA NO RECONHECIMENTO
E CLASSIFICAÇÃO DE NASCIDOS VIVOS QUANTO À SOBREVIVÊNCIA OU
ÓBITO NO PRIMEIRO ANO DE VIDA**

Dissertação de Mestrado apresentada como requisito parcial para a obtenção do grau de Mestre em Ciências, Área de Concentração em Programação Matemática, do Programa de Pós-Graduação em Métodos Numéricos em Engenharia dos departamentos de Matemática e de Construção Civil da Universidade Federal do Paraná, sob orientação do Professor Dr. Anselmo Chaves Neto.

CURITIBA

2006

IVAN DE OLIVEIRA

**FUNÇÃO DISCRIMINANTE QUADRÁTICA APLICADA NO RECONHECIMENTO
E CLASSIFICAÇÃO DE NASCIDOS VIVOS QUANTO À SOBREVIVÊNCIA OU
ÓBITO NO PRIMEIRO ANO DE VIDA**

Dissertação aprovada como requisito parcial para a obtenção do grau de Mestre em Ciências, Área de Concentração em Programação Matemática, na Linha de Pesquisa em Métodos Estatísticos Aplicados à Engenharia do Programa de Pós-Graduação em Métodos Numéricos em Engenharia dos departamentos de Matemática e de Construção Civil da Universidade Federal do Paraná, pela comissão examinadora formada pelos professores:

Orientador: Prof. Anselmo Chaves Neto, D.Sc.
Programa de Pós-graduação em Métodos Numéricos em Engenharia e DEST
Universidade Federal do Paraná

Profª. Maria Teresinha Arns Steiner, Dra. Eng.
Programa de Pós-graduação em Métodos Numéricos em Engenharia e DMAT
Universidade Federal do Paraná

Prof. Júlio César Nievola, Dr. Eng.
Programa de Pós Graduação em Informática Aplicada
Pontifícia Universidade Católica do Paraná

Curitiba, 25 de agosto de 2006.

AGRADECIMENTO

À Deus, pela inspiração na realização deste trabalho;

Aos meus pais, pela minha vida e pelo caminho que me ensinaram a seguir;

Ao professor e amigo Anselmo Chaves Neto, pela ajuda, atenção, paciência e orientação do trabalho;

Aos amigos Carlos Domanski, George Maluf e Cleverson Bussolo Kletemberg, pela colaboração no desenvolvimento deste trabalho;

À todos os amigos do CESEC, pelo companheirismo;

À 6ª Regional de Saúde e a Vigilância Epidemiológica de União da Vitória, pela colaboração e acesso aos dados necessários à elaboração do presente estudo.

Dedico este trabalho à minha esposa, pelo seu apoio incondicional mesmo nas horas mais difíceis, permitindo-me chegar ao fim de mais esta etapa de minha vida.

“Pode-se dizer a meu respeito que apenas montei um buquê com as flores de outros homens, e nada trouxe de meu próprio a não ser o cordão que as une, o qual lhe ofereço com prazer”.

MICHEL SEIGNEUR DE MONTAIGNE

SUMÁRIO

LISTA DE TABELAS	ix
LISTA DE FIGURAS.....	x
LISTA DE SIGLAS E ABREVIATURAS	xiii
RESUMO.....	xiv
ABSTRACT	xv
INTRODUÇÃO	1
1.1 PROBLEMA	1
1.2 OBJETIVOS.....	3
1.2.1 Objetivo geral	3
1.2.2 Objetivos específicos.....	3
1.3 JUSTIFICATIVA	3
1.4 ESTRUTURA DO TRABALHO	4
CAPÍTULO II.....	5
2 REVISÃO BIBLIOGRÁFICA	5
2.1 ANÁLISE MULTIVARIADA	8
2.2 ANÁLISE DISCRIMINANTE	10
2.2.1 O problema geral de reconhecimento e classificação.....	12
2.2.2 Regiões de classificação para duas populações.....	14
2.2.3 Matriz do custo de reconhecimento (classificação) errado e ECM.....	15
2.2.4 Critério TPM	18
2.3 CLASSIFICAÇÃO COM DUAS POPULAÇÕES NORMAIS MULTIVARIADAS	19
2.3.1 Classificação com duas populações normais multivariadas: covariâncias iguais: $\Sigma_1 = \Sigma_2 = \Sigma$	19
2.4 FUNÇÃO DISCRIMINANTE QUADRÁTICA.....	23
2.4.1 Classificação quadrática: populações com covariâncias diferentes: $\Sigma_1 \neq \Sigma_2$	23
2.5 FUNÇÃO DISCRIMINANTE LINEAR DE FISHER	27
2.5.1 Função discriminante linear para duas populações	28
2.6 REGRESSÃO LOGÍSTICA.....	30
2.6.1 Regressão logística para resposta dicotômica	31

2.6.2 Modelo de regressão logístico simples.....	33
2.6.3 Modelo de regressão logística múltiplo.....	36
2.6.4 Métodos de estimação dos parâmetros por máxima verossimilhança (MV).....	37
2.6.4.1 Método do gradiente.....	37
2.6.4.2 Método de Newton	37
2.6.4.3 Método de Levenberg-Marquardt.....	38
2.7 AVALIAÇÃO DA EFICIÊNCIA	40
2.8 A DISTRIBUIÇÃO NORMAL.....	41
2.8.1 Distribuição normal multivariada.....	41
2.8.2 Avaliando a normalidade bivariada.....	44
2.8.3 Avaliando a normalidade de uma distribuição com $p \geq 2$	44
2.9 MORTALIDADE INFANTIL: FATORES E INDICADORES	45
2.9.1 A mortalidade infantil no mundo	46
2.9.2 A mortalidade infantil no Brasil	47
2.9.2.1 Principais causas da mortalidade infantil	49
2.9.3 Índice de Apgar	51
2.9.4 Coeficiente de mortalidade infantil	52
2.9.5 Declaração de nascido vivo	54
CAPÍTULO III	57
3 MATERIAIS E MÉTODOS.....	57
3.1 DADOS DO ESTUDO.....	57
3.2 VARIÁVEIS ANALISADAS	58
3.3 PROCESSAMENTO DOS DADOS	59
CAPÍTULO IV.....	61
4 RESULTADOS E DISCUSSÃO	61
4.1 ANÁLISES PRELIMINARES.....	61
4.2 AVALIAÇÃO DA NORMALIDADE DOS DADOS	61
4.3 ANÁLISE DO MODELO	63
4.4 AVALIAÇÃO DO MODELO	85
4.4.1 Análise dos resultados sem exclusão de variáveis	85
4.4.2 Análise dos resultados com exclusão de variáveis	86

CONCLUSÕES.....	91
REFERÊNCIAS	93
ANEXO I – DECLARAÇÃO DE NASCIDO VIVO (DN).....	96
ANEXO II – DECLARAÇÃO DE ÓBITO (DO).....	97
ANEXO III – CÓDIGO FONTE DO PROGRAMA.....	98
ANEXO IV – MANUAL DO PROGRAMA	114

LISTA DE TABELAS

TABELA 2.1 – TAXA DE MORTALIDADE INFANTIL - BRASIL, GRANDES REGIÕES E ESTADOS: 1980 - 2004.....	48
TABELA 4.1 – ANÁLISE DOS RESULTADOS SEM EXCLUSÃO DE VARIÁVEIS – PROBABILIDADE A PRIORI DIFERENTES	85
TABELA 4.2 – ANÁLISE DOS RESULTADOS SEM EXCLUSÃO DE VARIÁVEIS - PROBABILIDADES A PRIORI IGUAIS	86
TABELA 4.3 – ANÁLISE DOS RESULTADOS EXCLUINDO AS VARIÁVEIS UMA A UMA...	86
TABELA 4.4 – ANÁLISE DOS RESULTADOS UTILIZANDO AS TRÊS VARIÁVEIS CONSIDERADAS IMPORTANTES PELA LITERATURA MÉDICA.....	87
TABELA 4.5 – ANÁLISE DOS RESULTADOS UTILIZANDO AS QUATRO VARIÁVEIS CONSIDERADAS IMPORTANTES PELA LITERATURA MÉDICA.....	88
TABELA 4.6 – ANÁLISE DOS RESULTADOS EXCLUINDO AS VARIÁVEIS UMA A UMA...	89
TABELA 4.7 – ANÁLISE DOS RESULTADOS UTILIZANDO AS TRÊS VARIÁVEIS CONSIDERADAS IMPORTANTES PELA LITERATURA MÉDICA.....	89
TABELA 4.8 – ANÁLISE DOS RESULTADOS UTILIZANDO AS QUATRO VARIÁVEIS CONSIDERADAS IMPORTANTES PELA LITERATURA MÉDICA.....	90

LISTA DE FIGURAS

FIGURA 2.1 – DIAGRAMA DE DISPERSÃO	12
FIGURA 2.2 – REGIÕES DE CLASSIFICAÇÃO PARA DUAS POPULAÇÕES	13
FIGURA 2.3 – SEPARAÇÃO E CLASSIFICAÇÃO PARA AS DUAS POPULAÇÕES – SOBREVIVENTES E NÃO SOBREVIVENTES.	14
FIGURA 2.4 – MATRIZ DO CUSTO DE RECONHECIMENTO ERRADO E ECM	15
FIGURA 2.5 – GRÁFICO DA FUNÇÃO SIGMÓIDE	32
FIGURA 4.1 – TESTE DE NORMALIDADE PARA O GRUPO SOBREVIVENTE	62
FIGURA 4.2 – TESTE DE NORMALIDADE PARA O GRUPO NÃO-SOBREVIVENTE	62
FIGURA 4.3 – RESULTADOS COM AMOSTRA 01 – PROBABILIDADE A PRIORI IGUAIS	63
FIGURA 4.4 – RESULTADOS COM AMOSTRA 02 – PROBABILIDADE A PRIORI IGUAIS	63
FIGURA 4.5 – RESULTADOS COM AMOSTRA 03 – PROBABILIDADE A PRIORI IGUAIS	64
FIGURA 4.6 – RESULTADOS COM AMOSTRA 04 – PROBABILIDADE A PRIORI IGUAIS	64
FIGURA 4.7 – RESULTADOS COM AMOSTRA 05 – PROBABILIDADE A PRIORI IGUAIS	65
FIGURA 4.8 – RESULTADOS COM UTILIZAÇÃO DE TODAS AS OBSERVAÇÕES – PROBABILIDADES A PRIORI IGUAIS	65
FIGURA 4.9 – RESULTADOS COM AMOSTRA 01 – PROBABILIDADE A PRIORI DIFERENTES.....	66
FIGURA 4.10 – RESULTADOS COM AMOSTRA 02 – PROBABILIDADE A PRIORI DIFERENTES.....	66
FIGURA 4.11 – RESULTADOS COM AMOSTRA 03 – PROBABILIDADE A PRIORI DIFERENTES.....	67
FIGURA 4.12 – RESULTADOS COM AMOSTRA 04 – PROBABILIDADE A PRIORI DIFERENTES.....	67
FIGURA 4.13 – RESULTADOS COM AMOSTRA 05 – PROBABILIDADE A PRIORI DIFERENTES.....	68
FIGURA 4.14 – RESULTADOS COM UTILIZAÇÃO DE TODAS AS OBSERVAÇÕES – PROBABILIDADES A PRIORI DIFERENTES	68
FIGURA 4.15 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL IDADE DA MÃE.....	69
FIGURA 4.16 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL ESTADO CIVIL DA MÃE	69
FIGURA 4.17 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL ESCOLARIDADE DA MÃE	70
FIGURA 4.18 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL DURAÇÃO	

DA GESTAÇÃO	70
FIGURA 4.19 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL TIPO DE GRAVIDEZ	71
FIGURA 4.20 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL TIPO DE PARTO	71
FIGURA 4.21 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL NUMERO DE CONSULTAS PRÉ-NATAL.....	72
FIGURA 4.22 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL SEXO	72
FIGURA 4.23 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL APGAR 1º MINUTO.....	73
FIGURA 4.24 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL APGAR 5º MINUTO.....	73
FIGURA 4.25 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL PESO.....	74
FIGURA 4.26 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL NÚMERO DE GESTAÇÕES ANTERIORES	74
FIGURA 4.27 – RESULTADOS OBTIDOS COM A UTILIZAÇÃO DAS VARIÁVEIS APGAR 1, APGAR 5 E PESO	75
FIGURA 4.28 – RESULTADOS OBTIDOS COM A EXCLUSÃO DAS VARIÁVEIS APGAR 1, APGAR 5 E PESO	75
FIGURA 4.29 – RESULTADOS OBTIDOS COM A UTILIZAÇÃO DAS VARIÁVEIS IDADE DA MÃE, APGAR 1, APGAR 5 E PESO.....	76
FIGURA 4.30 – RESULTADOS OBTIDOS COM A EXCLUSÃO DAS VARIÁVEIS IDADE DA MÃE, APGAR 1, APGAR 5 E PESO.....	76
FIGURA 4.31 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL IDADE DA MÃE.....	77
FIGURA 4.32 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL ESTADO CIVIL DA MÃE	77
FIGURA 4.33 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL ESCOLARIDADE DA MÃE	78
FIGURA 4.34 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL DURAÇÃO DA GESTAÇÃO	78
FIGURA 4.35 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL TIPO DE GRAVIDEZ.....	79
FIGURA 4.36 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL TIPO DE PARTO	79
FIGURA 4.37 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL NÚMERO	

DE CONSULTAS PRÉ-NATAL.....	80
FIGURA 4.38 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL SEXO	80
FIGURA 4.39 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL APGAR 1º	
MINUTO.....	81
FIGURA 4.40 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL APGAR 5.º	
MINUTO.....	81
FIGURA 4.41 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL PESO.....	82
FIGURA 4.42 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL NÚMERO	
DE GESTAÇÕES ANTERIORES	82
FIGURA 4.43 – RESULTADOS OBTIDOS COM A UTILIZAÇÃO DAS VARIÁVEIS APGAR 1,	
APGAR 5 E PESO	83
FIGURA 4.44 – RESULTADOS OBTIDOS COM A EXCLUSÃO DAS VARIÁVEIS APGAR 1,	
APGAR 5 E PESO	83
FIGURA 4.45 – RESULTADOS OBTIDOS COM A UTILIZAÇÃO DAS VARIÁVEIS IDADE	
DA MÃE, APGAR 1, APGAR 5 E PESO.....	84
FIGURA 4.46 – RESULTADOS OBTIDOS COM A EXCLUSÃO DAS VARIÁVEIS IDADE	
DA MÃE, APGAR 1, APGAR 5 E PESO.....	84

LISTA DE SIGLAS E ABREVIATURAS

ABREVIATURAS

DM - *Data Mining*

DN - Declaração de Nascimento

DO - Declaração de Óbito

MI - Mortalidade Infantil

MS - Ministério da Saúde

NV - Nascido Vivo

SIGLAS

IBGE - Instituto Brasileiro de Geografia e Estatística

KDD - *Knowledge Discovery in Databases*

SIM - Sistema de Informações sobre Mortalidade

SINASC - Sistema de Informações sobre Nascidos Vivos

RESUMO

O índice de mortalidade infantil representa um dos principais componentes utilizados pela saúde pública para avaliar a qualidade de vida da população. A crescente disponibilidade de informações contidas no Sistema de Informação de Nascidos Vivos e Sistema de Informações de Mortalidade tornou possível relacionar as diversas informações como o *status* financeiro, escolaridade dos pais, idade da mãe e peso da criança, entre outras, com o óbito ou não do recém-nascido. O objetivo deste estudo é o desenvolvimento de um modelo de identificação do nascido vivo em risco de óbito e a verificação da influência de diversas variáveis neste fato, a partir das informações contidas no Sistema de Informação de Nascidos Vivos. O modelo foi desenvolvido com base em técnicas de análise multivariada e possibilita a tomada de providências por parte dos agentes públicos de saúde. A aplicação da Função Discriminante Quadrática gerou uma regra de reconhecimento e conseqüente classificação que apresentou resultados corretos de classificação para 97,20% das observações. Este resultado foi obtido usando probabilidades a priori diferentes, 1,2% para não sobrevivência e 98,8% para sobrevivência, utilizando-se apenas as variáveis Apgar 1, Apgar 5 e peso do recém-nascido, citadas como importantes pela literatura médica na determinação ou não de óbito. Com probabilidades a priori iguais a 50%, o melhor resultado classificou corretamente 92,81% das observações, com o uso das mesmas variáveis. O melhor resultado para classificação correta de crianças não sobreviventes ao primeiro ano de vida foi de 63,27%, obtido com o uso de todas as variáveis do estudo.

Palavras-chaves: *Data Mining*; Análise Multivariada; Mortalidade Infantil.

ABSTRACT

The infantile mortality rate represents one of the main components used by public health to evaluate the population quality of life. The increasing availability of information contained in the System of Born Alive Information and the System of Mortality Information became possible to relate the different information such as economic status, parents education, mother age and child weight, among others, with the death or not of the newborn. The aim of this study is the development of an identification pattern for the born alive in risk of infantile death and the verification of the influence of different variables in this fact, from the information contained in the System of Born Alive Information. The pattern was developed through out multi-variant analysis techniques and that can allow taking measures by the public health agents. The applicability of the Discriminating Square Function conceived a recognizing and classification rule that presented correct classification results to 97,20% of the observations. This result was obtained using probabilities a priori different, 1,2% to non-survival and 98,8% to survival, using only the variable Apgar1, Apgar5 and the newborn weight, mentioned as important by the medical literature to determine the death or not. With the same a priori probabilities to 50%, the best result classified correctly 92,81% of the observations, with the use of the same variables. The best results to classify correctly non-surviving children during the first year of life was of 63,27%, obtained with the use of all variables from the study.

Keywords: Data Mining; Multivariate Analysis; Infantile Mortality.

INTRODUÇÃO

1.1 PROBLEMA

O desenvolvimento dos sistemas informatizados nas últimas décadas transformou profundamente quase todas as atividades humanas, convertendo-se em uma excepcional ferramenta de trabalho em campos tão diversos quanto as comunicações, o ensino, a medicina e as ciências em geral. Esse desenvolvimento tecnológico, oriundo das descobertas científicas e informatização, tem alavancado o desenvolvimento científico, ampliando a capacidade de se obter informações de acontecimentos e fenômenos variados.

O fato dos sistemas informatizados gerarem uma enorme quantidade de dados e, a partir delas, ser necessária a extração de informações específicas para previsões e tomadas de decisão, exigiu o desenvolvimento de ferramentas que pudessem tratar estas informações, e delas se extraírem tendências e padrões.

Estas ferramentas conduziram ao grande desenvolvimento de técnicas conhecidas como *Data Mining* ou Mineração de Dados, que juntamente com uma conjunção de fatores, como a grande quantidade e acúmulo de dados, assim como o relativo barateamento do processamento dos mesmos, tornou-se amplamente utilizado.

As aplicações são variadas, entre elas a análise de riscos, a determinação de perfis de clientes, a detecção de fraudes, dentre outras, podendo ser aplicadas nas mais variadas áreas de pesquisa, tais como: economia, medicina, política, educação, sistemas de informação, engenharia, administração, entre outros.

Para se adquirir conhecimento das bases de dados são utilizadas técnicas, de reconhecimento de padrões, de aprendizado de máquina e métodos de visualização para se obter regras que darão suporte à análise de dados, descobrindo conhecimentos embutidos nestes dados.

Quando um fenômeno depende de muitas variáveis não basta apenas conhecer informações estatísticas isoladas; é necessário também conhecer a totalidade destas informações fornecida pelo conjunto das variáveis. Assim, a Análise Multivariada de Dados é

útil, pois utiliza simultaneamente todas as variáveis na interpretação teórica do conjunto de dados obtidos [MOITA NETO, 2006].

O objetivo deste trabalho é a aplicação de técnicas de Análise Multivariada com o propósito de detectar como a interação entre os fatores biológicos, sócio-econômicos e assistenciais informados na Declaração de Nascidos Vivos é determinante da mortalidade durante o primeiro ano de vida, criando uma tendência de óbito (ou não) em recém-nascidos vivos na região que corresponde a 6ª Regional de Saúde. Assim, pode-se criar medidas que garantam uma gestação e nascimento em condições tais que favoreçam a sobrevivência do indivíduo.

Sabe-se que a redução nos índices de mortalidade infantil observada nas últimas décadas em todo o país, embora distribuída de maneira desigual entre a população, remete ao questionamento sobre quais os fatores que poderiam ser responsáveis por esse desfecho [CRUZ, 1996].

As técnicas de Análise Multivariada geralmente utilizadas para a obtenção deste tipo de informação são a Análise Discriminante e a Regressão Logística, pois, são as técnicas estatísticas apropriadas quando a variável é categórica (nominal ou não-métrica) e as variáveis independentes são métricas [HAIR *et. al*, 2005] e próprias para reconhecimento e classificação.

A Análise Discriminante e a Regressão Logística são utilizadas para a resolução de problemas que envolvem a separação de conjuntos distintos de observações e a alocação de novas observações em um conjunto específico.

Além disso, segundo CRUZ [1996], na utilização do modelo de Regressão Logística, pode-se estar interessado no efeito de um fator de risco específico ou na identificação de vários fatores associados com a variável resposta, que no caso dos estudos na área de saúde são em sua maioria binários.

A relevância deste estudo está no desenvolvimento de um modelo de identificação de risco de mortalidade infantil a partir das informações contidas no Sistema de Informação de Nascidos Vivos e Sistema de Informações de Mortalidade na cidade de União da Vitória, de forma propiciar, de forma abrangente e confiável, uma base para os diversos níveis de

gerenciamento público em suas ações de saúde.

1.2 OBJETIVOS

1.2.1 Objetivo geral

O objetivo geral deste estudo é verificar se as variáveis descritas como determinantes do óbito infantil na literatura médica são realmente importantes, evidenciando a probabilidade de óbito (ou não) em recém-nascidos durante o primeiro ano de vida.

Para tal fim, utilizar-se-á o banco de dados da 6ª Regional de Saúde e, através da utilização de técnica estatística multivariada denominada Função Discriminante Quadrática, prever o óbito (ou não) de uma criança no primeiro ano de vida a partir de sua Declaração de Nascido Vivo.

1.2.2 Objetivos específicos

- a) Identificar dados (padrões) e atributos relevantes na determinação do óbito (ou não) com base na literatura médica;
- b) Verificar se as variáveis relevantes descritas na literatura médica são determinantes do óbito ou não;
- c) Elaborar uma ferramenta computacional que possa reconhecer e classificar o recém-nascido, quanto à tendência a óbito (ou não).

1.3 JUSTIFICATIVA

Os fatores determinantes da mortalidade infantil são múltiplos e complexos, relacionando-se à interação de variáveis: sócio-econômicas, biológicas e assistenciais.

Como exemplo, as variáveis biológicas referem-se à mãe e ao recém-nascido e são as causas diretas dos óbitos em recém-nascidos. O acesso aos serviços de saúde e a qualidade da assistência prestada no pré-natal, sala de parto e os cuidados pós-natais ao recém-nascido

são variáveis assistenciais capazes de interferir nos fatores de riscos biológicos e sócio-econômicos para a mortalidade infantil. As variáveis sócio-econômicas indicam as condições em que vive a mãe, as quais são capazes de influenciar nas variáveis biológicas e dificultar ou facilitar o acesso a uma adequada assistência no período da gestação e nascimento.

Neste sentido, o presente estudo justifica-se pelo desafio que a redução da mortalidade infantil representa para governo e sociedade através da identificação de quais fatores lhe são determinantes. A identificação desses fatores fornecerá subsídios para a avaliação da situação da mortalidade durante o primeiro ano de vida, identificando pontos estratégicos para intervenções, reduzindo situações indesejadas.

1.4 ESTRUTURA DO TRABALHO

A metodologia utilizada para elaboração do presente estudo apresenta-se estruturado da seguinte forma: além desta introdução tem-se no segundo capítulo um estudo de cunho teórico bibliográfico, abordando os principais conceitos, definições, aspectos e aplicação da Análise Multivariada nos mais diversos campos de pesquisa, dando um suporte para a continuidade do estudo. Também são abordados aspectos relacionados à mortalidade infantil e formas de aferição da mesma. No terceiro capítulo é apresentada a pesquisa experimental, demonstrando os grupos de observações estudados, métodos e variáveis analisadas. O quarto capítulo apresenta os resultados obtidos com a ferramenta computacional implementada e discussão destes. O presente estudo é finalizado com as conclusões finais, avaliando e validando os resultados obtidos no modelo proposto, apresentando sugestões para trabalhos futuros.

CAPÍTULO II

2 REVISÃO BIBLIOGRÁFICA

A Estatística é a ciência que trata da coleta, da análise, da interpretação e da apresentação de massas de dados numéricos com ênfase principal na análise exploratória, sendo importante nas ciências aplicadas como *marketing*, economia, medicina, biologia, agricultura, educação, psicologia, sociologia, etc.

Primeiramente, o termo estatística foi utilizado no século XVIII para designar o conjunto de dados referentes a assuntos de interesse do estado, geralmente com finalidade de controle fiscal ou de segurança nacional e esses dados referiam-se, particularmente, à população, às transações comerciais internas ou com outros países, ao controle da mortalidade e aos problemas de taxaço e de proporcionalidade de tarifas e impostos [ZANDONADE, 2006].

Atualmente, a Estatística se caracteriza pelo aprimoramento técnico-científico, pelo aparecimento e desenvolvimento de novas teorias e associação com diversas áreas de conhecimento.

A Estatística hoje é uma ciência de apoio para diversos ramos do conhecimento. Na área biomédica, a Bioestatística atende desde a análise de dados clínicos, passando pela pesquisa científica e sua utilização em artigos especializados [ZANDONADE, 2006].

Por ser uma técnica auxiliar do estudo dos fenômenos coletivos, sejam eles econômicos, sociais ou científicos, a Estatística é um método de observação, descrição, mensuração e interpretação dos fenômenos coletivamente típicos e da indagação de suas uniformidades e relações.

LOPES [2006] considera que a Estatística é freqüentemente apresentada como um testemunho de credibilidade a um argumento ou a uma recomendação, fato que se pode comprovar ouvindo os meios de comunicação: o primeiro pensamento é acreditar na notícia como se fosse verdade absoluta.

Seu campo de aplicação é o dos fenômenos que apresentam regularidade na

observação de massas de casos, embora uma parte de seus processos também encontre aplicação no domínio dos fenômenos atípicos e sua teoria permita que sejam tomadas decisões com base em informações geralmente limitadas e incompletas sobre os mais variados fenômenos que ocorrem no mundo [BARSA, 2001].

Assim, a Estatística moderna é uma tecnologia quantitativa para a ciência experimental e observacional que permite avaliar e estudar as incertezas e os seus efeitos no planejamento e interpretação de experiências e de observações de fenômenos da natureza e da sociedade [MATSUSHITA *apud* RAMOS, 2006].

A crescente utilização de sistemas informatizados nas áreas governamentais, corporativas e científicas e o crescimento dos bancos de dados, superando em muito a usual capacidade de interpretar e examinar estes dados, gerou a necessidade de novas ferramentas e técnicas para análise automática e inteligente de bancos de dados. Diante desta necessidade surgiu a área de estudo denominada Descoberta de Conhecimento em Bases de Dados (*Knowledge Discovery in Databases* - KDD) que pode ser definida como o processo de extração de informação útil, não trivial e previamente desconhecida da base de dados [ALMEIDA, 2003].

Segundo MULLER [2002], o termo *KDD* foi criado em 1989 para se referir ao amplo processo de descoberta de informação em dados. Isto pode ser entendido como o processo da descoberta de novas correlações, padrões e tendências significativas, por meio da análise minuciosa de grandes conjuntos de dados armazenados.

A mineração de dados caracteriza-se como uma atividade multidisciplinar, envolvendo principalmente as áreas de Banco de Dados (recursos para manipular grandes bases de dados), Inteligência Artificial (especialmente aprendizagem de máquina) e Estatística (comumente na avaliação e validação de resultados), além de envolver conhecimentos específicos sobre os dados a serem minerados e os objetivos da entidade interessada na mineração.

Este processo se vale de tecnologias de reconhecimento utilizando padrões e técnicas estatísticas e matemáticas e incluem investigação e criação de conhecimento, processos, algoritmos e mecanismos de recuperação de conhecimento potencial em dados

armazenados.

Portanto, a Descoberta de Conhecimento em Bases de Dados está associada à identificação de relacionamentos implícitos existentes nos dados analisados e seu processo se desenrola em várias fases que inclui a gestão dos algoritmos de *Data Mining*, utilizados para extrair padrões dos dados, e a interpretação dos padrões encontrados pelos mesmos [CARVALHO, *et. al.*, 2003].

Data Mining é formado por um conjunto de ferramentas que através do uso de algoritmos de aprendizado ou baseados em redes neurais e estatística, são capazes de explorar um grande conjunto de dados extraído, deste, conhecimento na forma de hipóteses e de regras.

Esta análise inteligente e automática de dados para descobrir padrões ou regularidades em grandes conjuntos de dados envolve métodos matemáticos, algoritmos baseados em conceitos biológicos, processos lingüísticos e heurística [QUISPER, 2003].

Os pesquisadores da área de inteligência artificial e estatística, além dos físicos que trabalham no domínio de dinâmicas não lineares contribuíram para o desenvolvimento de novos conjuntos de métodos lógicos. Entretanto, estes métodos exigem máquinas de alto desempenho e, por isto, somente recentemente puderam ser implementados, despertando interesse também de pesquisadores de outras áreas. A garimpagem de dados utiliza esses métodos, para que, a partir de um conjunto de dados, seja possível descobrir uma representação otimizada da sua estrutura.

O ponto central de um sistema de descoberta de conhecimento é formado pelos algoritmos de extração de padrões que, quando devidamente aplicados sobre bases de dados, auxiliam nas tarefas de mineração de dados. Os padrões que são detectados são associações interessantes entre os elementos contidos na base de dados, e podem ser de diferentes tipos.

Para ALMEIDA [2003] a fase mais importante do *KDD* é a Mineração de Dados e é nesta fase que informação útil é efetivamente extraída dos dados que pode utilizar várias abordagens.

Uma grande massa de informação deve ser processada antes de ser transformada em conhecimento. Portanto, cada vez mais estamos necessitando de ferramentas estatísticas

que apresentem uma visão mais global do fenômeno que aquela possível numa abordagem univariada.

2.1 ANÁLISE MULTIVARIADA

A análise multivariada vem a ser a análise estatística de múltiplas variáveis em um único relacionamento ou em um conjunto de relações. A rigor reúne um conjunto de técnicas para o estudo da estrutura de covariância de um vetor aleatório, do agrupamento de itens representados vetorialmente, para o reconhecimento e classificação de padrões, e outras.

Segundo HAIR *et. al.* [2005], a análise multivariada tem as suas raízes nas estatísticas univariada e bivariada e sua extensão multivariada introduz conceitos adicionais e questões que tem particular relevância.

Do ponto de vista da Estatística, as técnicas úteis para reconhecimento e descoberta de padrões em ambientes onde os fenômenos são descritos/baseados em uma grande variedade de dados são conhecidas como Análise de Dados.

A Análise de Dados está relacionada de forma próxima com o conceito de Mineração de Dados. De um ponto de vista das técnicas estatísticas, é utilizada para a identificação de relacionamentos sistemáticos entre variáveis quando não existem expectativas a priori acerca da natureza destes relacionamentos ou estas são incompletas. Em um processo de Análise de Dados típico, muitas variáveis diferentes são consideradas e comparadas. Isto é realizado utilizando-se uma grande variedade de técnicas e modelos matemáticos com o objetivo de se encontrar padrões nestes dados.

A análise de dados é aplicada ao reconhecimento e classificação de padrões de duas formas diferentes [VON WANGEIHEIM, 2006]:

- As técnicas de Análise de Dados já são em si técnicas de reconhecimento de padrões, já que são projetadas para detectar regularidades, correlações e fatores agrupadores ou diferenciadores em um conjunto de dados. Sob esta ótica, realizar mineração de dados com Análise de Dados ou aplicá-las a um problema, já é uma forma de realizar reconhecimento de padrões.
- As técnicas da Análise de Dados são utilizadas como ferramentas para a extração

de informações de conjuntos de dados com o objetivo de utilizar estas informações para a implementação de um classificador. Neste enfoque, o desenvolvimento completo de uma solução envolve mais do que as técnicas de Análise de Dados em si, pois exige que se utilize ainda outra técnica adicional para a implementação deste classificador.

Para o Reconhecimento de Padrões, os resultados da aplicação de técnicas da estatística exploratória a conjuntos de padrões ou a um fenômeno esperado, podem ser utilizados também para o desenvolvimento de classificadores que poderão então, a posteriori, serem utilizados para classificar novos dados produzidos pelo mesmo fenômeno anteriormente analisado através das técnicas de Análise de Dados.

O resultado dessa análise inicial através de técnicas exploratórias pode nos prover dados para a elaboração de um mecanismo de classificação utilizando técnicas tradicionais, através de informações sobre distribuições de dados ou variáveis-chaves para classificação dos dados em classes. Em casos onde não conhecemos a priori em quais e quantas classes os dados se permitem agrupar, pode-se inclusive determinar estas classes e utilizar esta informação para um posterior mecanismo de classificação.

Segundo VON WANGELHEIM [2006], a utilização de técnicas estatísticas para o desenvolvimento de classificadores em Reconhecimento de Padrões é, portanto, baseada em uma filosofia de dois passos:

- Análise de Dados;
- Escolha da técnica de Reconhecimento de Padrões;

A Análise de Dados é realizada a fim de gerar um conjunto de informações a partir de um conjunto inicial de dados gerado por um processo que desejamos dominar, capaz de servir para utilização em um classificador e subsequentemente, ser utilizado para a classificação de novos casos gerados pelo mesmo processo que gerou os dados originais.

A técnica de Reconhecimento de Padrões adequada ao tipo de informação gerada pelo método de Análise de Dados utilizado e também adequada ao tipo de classificação que queremos obter para dados futuros e utiliza-se a informação gerada pela Análise de Dados para alimentar ou implementar o classificador.

Segundo HAIR *et. al.* [2005], os tipos de técnicas multivariadas são:

- Análise de componentes principais e análise dos fatores comuns;
- Regressão múltipla;
- Análise discriminante múltipla;
- Análise multivariada de variância e covariância;
- Análise conjunta;
- Correlação canônica;
- Análise de agrupamentos;
- Escalonamento multidimensional;
- Análise de correspondência;
- Modelos lineares de correspondência;
- Modelagem de equações estruturais.

Alguns autores ainda classificam as Redes Neurais como técnicas da análise de dados, porém, com certeza, Redes Neurais compõem, junto com técnicas estatísticas e de programação matemática, o conjunto de técnicas de reconhecimento e classificação de padrões.

2.2 ANÁLISE DISCRIMINANTE

Diferentes métodos quantitativos como análise discriminante, regressão logística e, mais recentemente, técnicas de programação matemática, redes neurais e outras, são empregados na área de reconhecimento de padrões onde o grupo ou a classificação, a que os casos ou observações pertencem, são conhecidos.

A análise discriminante é uma técnica estatística que permite estudar diferenças entre dois ou mais grupos, em função de um conjunto de informações conhecidas para todos os elementos dos grupos.

É uma técnica estatística multivariada usada na resolução de problemas que envolvem a separação de conjuntos distintos de objetos, ou observações, e a alocação de novos objetos, ou observações em um grupo específico. O reconhecimento de padrões está presente em áreas como:

- Classificação de empresas;
- Processamento de sinais;
- Análise de sinais eletrocardiográficos;
- Reconhecimento de impressões digitais;
- Elaboração de perfis de consumidores;
- Diagnóstico médico preliminar.

Um dos objetivos da Análise Discriminante é determinar a que grupo, dentre dois ou mais definidos a priori, pertence um novo elemento, com base em várias características observadas para o mesmo. Cada característica constitui uma variável independente, contribuindo para a classificação [GUIMARÃES e CHAVES NETO, 2002].

A Análise Discriminante combina estas variáveis em uma ou mais funções, de modo a determinar, para cada elemento, escores de classificação. Estas funções são construídas de modo que os escores dos elementos de cada grupo se concentrem em torno do escore médio do grupo, fazendo com que a superposição de escores de elementos de diferentes grupos seja minimizada.

Segundo HUBERTY [*apud* SCARPEL, 2005], a análise discriminante implica na estimativa das densidades de probabilidades específicas nas diferentes populações. Em relação à escolha da função discriminante, ela pode depender do conhecimento prévio dos padrões que serão utilizados no processo de classificação ou pode-se optar por utilizar uma forma funcional específica com parâmetros estimados utilizando o conjunto de treinamento.

Como objetivos primordiais da utilização da discriminação e classificação, podem ser considerados:

- Dar uma descrição algébrica, ou gráfica, de características diferenciais das observações, com valores numéricos tais que permitam a máxima separação das populações estudadas.
- Encontrar uma regra que permita a alocação ótima de uma nova observação em um dentre dois ou mais grupos, de acordo com os critérios estabelecidos pela função encontrada.

Uma regra que permita a separação também possibilita a classificação, o que acaba

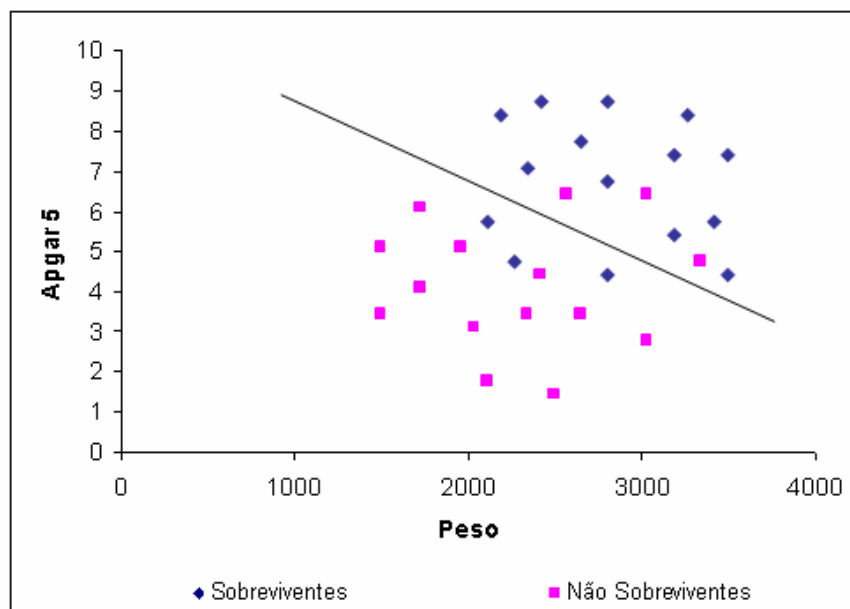
por gerar uma sobreposição dos objetivos anteriores.

A análise discriminante permite descobrir as ligações que existem entre um caráter qualitativo a ser explicado e um conjunto de caracteres quantitativos explicativos e permitem também prever, por meio de um modelo, as modalidades do caráter a ser explicado a partir dos valores tomados pelos caracteres explicativos.

2.2.1 O problema geral de reconhecimento e classificação

Considere dois grupos de crianças em um município: sobreviventes ao primeiro ano de vida e não sobreviventes. A fim de diminuir a taxa de mortalidade infantil no primeiro ano de vida, a vigilância epidemiológica do município está interessada em classificar crianças como prováveis sobreviventes ou não, com base em $X_1 = \text{Peso}$ e $X_2 = \text{Apgar no 5º minuto}$. Amostras aleatórias de $n_1 = 15$ crianças sobreviventes e $n_2 = 15$ crianças não sobreviventes ao primeiro ano de vida produziram o diagrama de dispersão da figura 2.1:

FIGURA 2.1 – DIAGRAMA DE DISPERSÃO



FONTE: O AUTOR

Observa-se que:

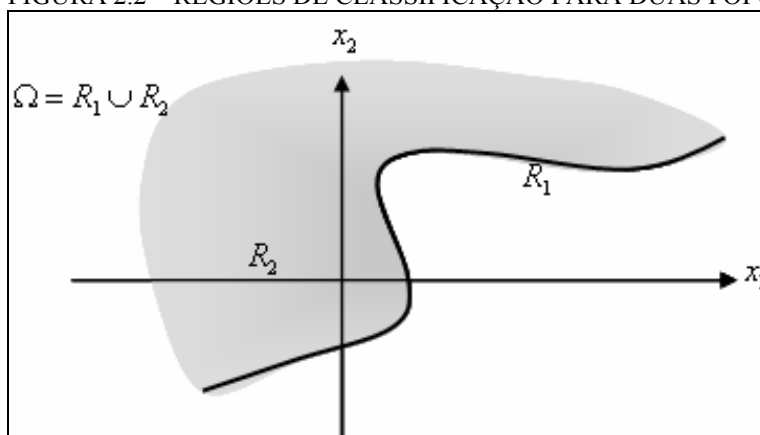
- 1) Crianças sobreviventes tendem a ter maiores pesos e valores para Apgar no 5º Minuto;

- 2) Apgar no 5º Minuto parece discriminar melhor do que Peso;
- 3) Existe “mistura” entre os grupos.

Como existe mistura e, conseqüentemente, classificações erradas, a idéia é criar uma regra (regiões R_1 e R_2) que minimize a chance de ocorrer esta mistura. Bons procedimentos de classificação resultarão em pouca mistura de elementos grupais. Uma regra de classificação ótima deve levar em conta as probabilidades de ocorrência a priori. Também se deve levar em conta o custo de uma classificação errada. Suponha que classificar um item em Π_1 quando na verdade ele pertence a Π_2 represente um erro mais sério do que classificá-lo em Π_2 quando ele pertence a Π_1 . Então, deve-se levar isto em conta.

Sejam $f_1(\underline{x})$ e $f_2(\underline{x})$ as f.d.p's associadas com o vetor aleatório \underline{X} de dimensão p das populações Π_1 e Π_2 , respectivamente. Um objeto, com as medidas \underline{X} , deve ser reconhecido como de Π_1 ou de Π_2 . Seja Ω o espaço amostral, isto é, o conjunto de todas as possíveis observações de \underline{X} . Seja R_1 o conjunto de valores \underline{X} para os quais se classifica o objeto como de Π_1 e $R_2 = \Omega - R_1$ os remanescentes valores \underline{X} para os quais se classifica os objetos como Π_2 . Os conjuntos R_1 e R_2 são mutuamente exclusivos. Para $p = 2$, podemos ter a figura 2.2:

FIGURA 2.2 – REGIÕES DE CLASSIFICAÇÃO PARA DUAS POPULAÇÕES



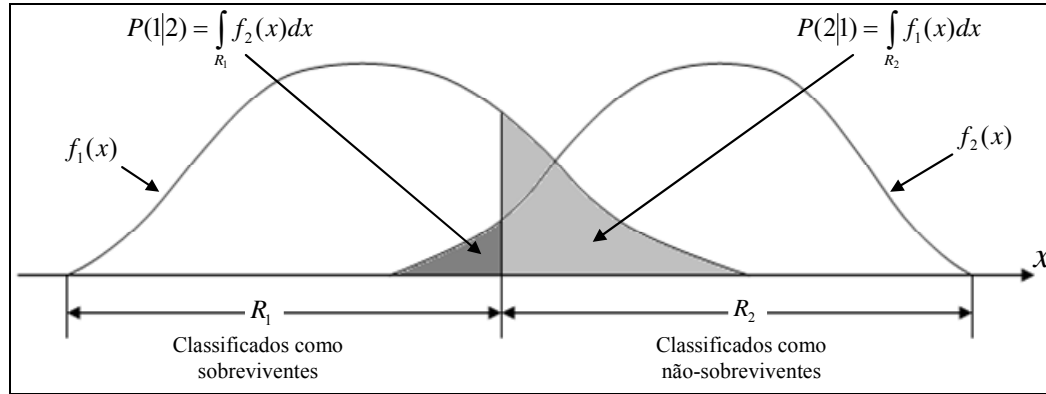
FONTE: O AUTOR

2.2.2 Regiões de classificação para duas populações

A probabilidade condicional de reconhecer um objeto como de Π_2 quando na verdade ele é de Π_1 é: $P(2|1) = P(\underline{X} \in R_2 | \Pi_1) = \int_{R_2 = \Omega - R_1} f_1(\underline{x}) d\underline{x}$. Da mesma forma, a probabilidade condicional de reconhecer um objeto como de Π_1 quando na verdade ele é de Π_2 é: $P(1|2) = P(\underline{X} \in R_1 | \Pi_2) = \int_{R_1} f_2(\underline{x}) d\underline{x}$.

Segundo JOHNSON e WICHERN [1998], a $P(2|1)$ representa o volume formado pela f.d.p. $f_1(\underline{x})$ na região R_2 e sendo $p = 1$ (caso univariado) tem-se a situação da figura 2.3:

FIGURA 2.3 – SEPARAÇÃO E CLASSIFICAÇÃO PARA AS DUAS POPULAÇÕES – SOBREVIVENTES E NÃO SOBREVIVENTES.



FONTE: O AUTOR

Seja p_1 a probabilidade a priori de Π_1 e p_2 a probabilidade a priori de Π_2 , onde $p_1 + p_2 = 1$. As probabilidades de reconhecer corretamente ou incorretamente são dadas por:

$$P(\text{reconhecer corretamente como } \Pi_1) = P(\underline{X} \in \Pi_1 \text{ e é reconhecido corretamente como } \Pi_1)$$

$$= P(\underline{X} \in R_1 | \Pi_1) P(\Pi_1) = P(1|1) p_1$$

$$P(\text{reconhecer incorretamente como } \Pi_1) = P(\underline{X} \in \Pi_2 \text{ e é reconhecido incorretamente como } \Pi_1)$$

$$= P(\underline{X} \in R_1 | \Pi_2) P(\Pi_2) = P(1|2) p_2$$

$$P(\text{reconhecer corretamente como } \Pi_2) = P(\underline{X} \in \Pi_2 \text{ e é reconhecido corretamente como } \Pi_2)$$

$$= P(\underline{X} \in R_2 | \Pi_2) P(\Pi_2) = P(2|2) p_2$$

$$P(\text{reconhecer incorretamente como } \Pi_2) = P(\underline{X} \in \Pi_1 \text{ e é reconhecido incorretamente como } \Pi_2)$$

$$= P(\underline{X} \in R_2 | \Pi_1) P(\Pi_1) = P(2|1) p_1$$

2.2.3 Matriz do custo de reconhecimento (classificação) errado e ECM

Regras de reconhecimento e classificação são freqüentemente avaliadas em termos de suas probabilidades de reconhecimento e classificação erradas. Assim, é comum construir a matriz a seguir, ilustrada na figura 2.4, denominada de “Matriz de Confusão”.

FIGURA 2.4 – MATRIZ DO CUSTO DE RECONHECIMENTO ERRADO E ECM

População verdadeira		Reconhecido como	
		Π_1	Π_2
	Π_1	0	$c(2 1)$
	Π_2	$c(1 2)$	0

FONTE: O AUTOR

Para qualquer regra, a média ou o custo esperado de reconhecimento (classificação) errado é dado pela soma dos produtos dos elementos fora da diagonal principal pelas respectivas probabilidades:

$$ECM = c(2|1)p(2|1)p_1 + c(1|2)p(1|2)p_2$$

Uma regra razoável de reconhecimento deve ter ECM (*Expected Cost of Misclassification*) mínimo, tanto quanto possível.

Resultado 1:

As regiões R_1 e R_2 que minimizam o ECM são definidas pelos valores de \underline{x} tal que valem as desigualdades:

$$R_1 : \frac{f_1(\underline{x})}{f_2(\underline{x})} \geq \left[\frac{c(1|2)}{c(2|1)} \right] \left[\frac{p_2}{p_1} \right]$$

$$\left[\begin{array}{c} \text{Razão das} \\ \text{densidades} \end{array} \right] \geq \left[\begin{array}{c} \text{Razão dos} \\ \text{custos} \end{array} \right] \left[\begin{array}{c} \text{Razão das} \\ \text{probabilidades 'a priori'} \end{array} \right]$$

$$R_2 : \frac{f_1(\underline{x})}{f_2(\underline{x})} < \left[\frac{c(1|2)}{c(2|1)} \right] \left[\frac{p_2}{p_1} \right]$$

$$\left[\begin{array}{c} \text{Razão das} \\ \text{densidades} \end{array} \right] < \left[\begin{array}{c} \text{Razão dos} \\ \text{custos} \end{array} \right] \left[\begin{array}{c} \text{Razão das} \\ \text{probabilidades 'a priori'} \end{array} \right]$$

Prova:

$$ECM = c(2|1)P(2|1)p_1 + c(1|2)P(1|2)p_2$$

$$ECM = c(2|1)p_1 \int_{R_2} f_1(\underline{x})d\underline{x} + c(1|2)p_2 \int_{R_1} f_2(\underline{x})d\underline{x}$$

e como $\Omega = R_1 \cup R_2$ tem-se:

$$1 = \int_{\Omega} f_1(\underline{x})d\underline{x} = \int_{R_1} f_1(\underline{x})d\underline{x} + \int_{R_2} f_1(\underline{x})d\underline{x}$$

podemos escrever:

$$ECM = c(2|1)p_1 \left[1 - \int_{R_1} f_1(\underline{x})d\underline{x} \right] + c(1|2)p_2 \int_{R_1} f_2(\underline{x})d\underline{x}$$

e das propriedades de integral (volume).

$$ECM = \int_{R_1} [c(1|2)p_2 f_2(\underline{x}) - c(2|1)p_1 f_1(\underline{x})]d\underline{x} + c(2|1)p_1$$

onde p_1 , p_2 , $c(1|2)$ e $c(2|1)$ são não-negativos, e ainda $f_1(\underline{x})$ e $f_2(\underline{x})$ também são não-negativos e são as únicas quantidades de ECM que dependem de \underline{x} . Assim ECM é minimizado se R_1 inclui esses valores \underline{x} tal que $[c(1|2)p_2 f_2(\underline{x}) - c(2|1)p_1 f_1(\underline{x})] \leq 0$ e exclui aqueles para os quais esta quantidade é positiva. Isto é, R_1 deve ser o conjunto de pontos tal que:

$$c(1|2)p_2 f_2(\underline{x}) \leq c(2|1)p_1 f_1(\underline{x}) \Rightarrow \frac{f_1(\underline{x})}{f_2(\underline{x})} \geq \left[\frac{c(1|2)}{c(2|1)} \right] \left[\frac{p_2}{p_1} \right]$$

e dado que R_2 é o complemento de R_1 em Ω , R_2 deve ser o conjunto de pontos \underline{x} para os quais:

$$\frac{f_1(\underline{x})}{f_2(\underline{x})} < \left[\frac{c(1|2)}{c(2|1)} \right] \left[\frac{p_2}{p_1} \right]$$

Casos Especiais de Regiões de ECM:

a) Probabilidades a priori iguais: $\frac{p_2}{p_1} = 1$

$$R_1 = \frac{f_1(\underline{x})}{f_2(\underline{x})} \geq \frac{c(1|2)}{c(2|1)}; \quad R_2 = \frac{f_1(\underline{x})}{f_2(\underline{x})} < \frac{c(1|2)}{c(2|1)}$$

b) Custos de erro de reconhecimento iguais: $\frac{c(1|2)}{c(2|1)} = 1$

$$R_1 = \frac{f_1(\underline{x})}{f_2(\underline{x})} \geq \frac{p_2}{p_1}; \quad R_2 = \frac{f_1(\underline{x})}{f_2(\underline{x})} < \frac{p_2}{p_1}$$

c) Probabilidades a priori iguais e custos de reconhecimento errado iguais

$$\frac{p_2}{p_1} = \frac{c(1|2)}{c(2|1)} = 1 \text{ ou } \frac{p_2}{p_1} = \frac{1}{c(1|2)/c(2|1)}$$

$$R_1 = \frac{f_1(\underline{x})}{f_2(\underline{x})} \geq 1; \quad R_2 = \frac{f_1(\underline{x})}{f_2(\underline{x})} < 1$$

Observação:

1) Quando as probabilidades a priori são desconhecidas, elas são freqüentemente

tomadas como iguais e a razão de f.d.p's é comparada com a razão de custos de reconhecimento errado.

- 2) Se a razão de custo de reconhecimento errado é indeterminada, ela é usualmente tomada como 1 e a razão das f.d.p's é comparada com a razão de probabilidades a priori.
- 3) Finalmente, quando ambas: razão das probabilidades a priori e razão de custos são unitários ou uma razão é recíproca da outra, então as regiões de reconhecimento (classificação) ótimo são determinadas comparando-se os valores das f.d.p's. Assim, se \underline{x}_0 é uma nova observação e $f_1(\underline{x}_0)/f_2(\underline{x}_0) \geq 1 \Rightarrow f_1(\underline{x}_0) \geq f_2(\underline{x}_0)$, assume-se que $\underline{x}_0 \in \Pi_1$.

Avaliar o custo de reconhecimento errado quando se trata de mortalidade infantil é extremamente complexo. Pode-se assumir que não há custo algum em se classificar erroneamente uma criança que eventualmente permaneceria viva com uma criança que irá morrer.

Essa criança irá ter atendimento especial e isso apenas contribuirá ainda mais para que ela não morra. Agora, classificando-se erroneamente uma criança que chegaria a óbito no primeiro ano de vida como uma criança que permaneceria viva, ela deixaria de ter atendimento especial e isso contribuiria para que o óbito se efetivasse.

Assim o custo de reconhecimento errado na segunda situação pode ser considerado, socialmente, infinitamente maior que o custo de reconhecimento errado na primeira situação. Quantitativamente esses custos são indeterminados, consequentemente a razão de custos também e, segundo JOHNSON e WICHERN [1998], quando a razão dos custos de classificação é indeterminada é usual assumi-la como sendo unitária e então a razão de densidades populacionais é comparada com a razão de probabilidades a priori.

2.2.4 Critério TPM

Outro critério para a construção de regras de reconhecimento e classificação é escolher R_1 e R_2 que minimizem a probabilidade total de erro de classificação (*Total*

Productive Maintenance - TPM), dada por:

$$TPM = P(\underline{x} \in \Pi_1 \text{ e é classificada errada}) + P(\underline{x} \in \Pi_2 \text{ e é classificada errada})$$

$$TPM = p_1 \int_{R_2} f_1(\underline{x}) d\underline{x} + p_2 \int_{R_1} f_2(\underline{x}) d\underline{x}$$

Isto é equivalente a minimizar ECM quando os custos de classificação errada são iguais. Assim, pode-se alocar uma nova observação \underline{x}_0 para a população com a maior probabilidade posteriori $P(\Pi_i | \underline{x}_0)$, onde:

$$\begin{aligned} P(\Pi_1 | \underline{x}_0) &= \frac{P(\Pi_1 \text{ ocorre e observa-se } \underline{x}_0)}{P(\text{observa-se } \underline{x}_0)} \\ &= \frac{P(\text{observa-se } \underline{x}_0 | \Pi_1) P(\Pi_1)}{P(\text{observa-se } \underline{x}_0 | \Pi_1) p(\Pi_1) + P(\text{observa-se } \underline{x}_0 | \Pi_2) p(\Pi_2)} \\ &= \frac{p_1 f_1(\underline{x}_0)}{p_1 f_1(\underline{x}_0) + p_2 f_2(\underline{x}_0)} \end{aligned}$$

$$\text{e } P(\Pi_2 | \underline{x}_0) = 1 - P(\Pi_1 | \underline{x}_0) = \frac{p_2 f_2(\underline{x}_0)}{p_1 f_1(\underline{x}_0) + p_2 f_2(\underline{x}_0)}$$

e classifica-se \underline{x}_0 em Π_1 quando $P(\Pi_1 | \underline{x}_0) > P(\Pi_2 | \underline{x}_0)$

2.3 CLASSIFICAÇÃO COM DUAS POPULAÇÕES NORMAIS MULTIVARIADAS

2.3.1 Classificação com duas populações normais multivariadas: covariâncias iguais:

$$\Sigma_1 = \Sigma_2 = \Sigma$$

Considere $f_1(\underline{x})$ e $f_2(\underline{x})$ funções normais multivariadas:

- A primeira com vetor de médias $\underline{\mu}_1$ e matriz de covariância Σ_1 ;
- A segunda com vetor de médias $\underline{\mu}_2$ e matriz de covariância Σ_2 e que $\Sigma_1 = \Sigma_2 = \Sigma$.

A regra de classificação com base em ECM é a seguinte:

Resultado 2:

Se:

$$(\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} \underline{x} - \frac{1}{2} (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} (\underline{\mu}_1 + \underline{\mu}_2) \geq \ln \left[\left(\frac{c(1/2)}{c(2/1)} \right) \left(\frac{p_2}{p_1} \right) \right]$$

então \underline{x} pertence à região R_1 caso contrário pertence à região R_2 .

A primeira parcela da regra de reconhecimento e classificação, $(\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} \underline{x}$, é chamada Função Linear Discriminante de Fisher.

Prova:

Sabe-se que as regiões R_1 e R_2 que minimizam o ECM são definidas pelos valores de \underline{x} tal que valem as desigualdades:

$$R_1: \frac{f_1(\underline{x})}{f_2(\underline{x})} \geq \left[\frac{c(1/2)}{c(2/1)} \right] \left[\frac{p_2}{p_1} \right] \text{ e } R_2: \frac{f_1(\underline{x})}{f_2(\underline{x})} < \left[\frac{c(1/2)}{c(2/1)} \right] \left[\frac{p_2}{p_1} \right] \text{ onde}$$

$$\frac{f_1(\underline{x})}{f_2(\underline{x})} = \frac{\frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp \left[-\frac{1}{2} (\underline{x} - \underline{\mu}_1)' \Sigma^{-1} (\underline{x} - \underline{\mu}_1) \right]}{\frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp \left[-\frac{1}{2} (\underline{x} - \underline{\mu}_2)' \Sigma^{-1} (\underline{x} - \underline{\mu}_2) \right]}$$

Assim,

$$\frac{f_1(\underline{x})}{f_2(\underline{x})} = \frac{\exp \left[-\frac{1}{2} (\underline{x} - \underline{\mu}_1)' \Sigma^{-1} (\underline{x} - \underline{\mu}_1) \right]}{\exp \left[-\frac{1}{2} (\underline{x} - \underline{\mu}_2)' \Sigma^{-1} (\underline{x} - \underline{\mu}_2) \right]}$$

Aplicando o logaritmo natural obtém-se:

$$\ln\left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right) \geq \ln\left[\left(\frac{c(1/2)}{c(2/1)}\right)\left(\frac{p_2}{p_1}\right)\right]$$

Desenvolvendo o primeiro membro da desigualdade temos:

$$\begin{aligned} \ln\left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right) &= \ln\left(\exp\left[-\frac{1}{2}(\underline{x} - \underline{\mu}_1)' \Sigma^{-1}(\underline{x} - \underline{\mu}_1)\right]\right) - \ln\left(\exp\left[-\frac{1}{2}(\underline{x} - \underline{\mu}_2)' \Sigma^{-1}(\underline{x} - \underline{\mu}_2)\right]\right) \\ \ln\left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right) &= -\frac{1}{2}(\underline{x} - \underline{\mu}_1)' \Sigma^{-1}(\underline{x} - \underline{\mu}_1) + \frac{1}{2}(\underline{x} - \underline{\mu}_2)' \Sigma^{-1}(\underline{x} - \underline{\mu}_2) \\ \ln\left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right) &= -\frac{1}{2}(\underline{x}' \Sigma^{-1} \underline{x} - \underline{x}' \Sigma^{-1} \underline{\mu}_1 - \underline{\mu}_1' \Sigma^{-1} \underline{x} + \underline{\mu}_1' \Sigma^{-1} \underline{\mu}_1) \\ &\quad + \frac{1}{2}(\underline{x}' \Sigma^{-1} \underline{x} - \underline{x}' \Sigma^{-1} \underline{\mu}_2 - \underline{\mu}_2' \Sigma^{-1} \underline{x} + \underline{\mu}_2' \Sigma^{-1} \underline{\mu}_2) \end{aligned}$$

Como Σ^{-1} é matriz simétrica então:

$$\begin{aligned} \ln\left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right) &= -\frac{1}{2}(\underline{x}' \Sigma^{-1} \underline{x} - \underline{x}' \Sigma^{-1} \underline{\mu}_1 - \underline{x}' \Sigma^{-1} \underline{\mu}_1 + \underline{\mu}_1' \Sigma^{-1} \underline{\mu}_1) \\ &\quad + \frac{1}{2}(\underline{x}' \Sigma^{-1} \underline{x} - \underline{x}' \Sigma^{-1} \underline{\mu}_2 - \underline{x}' \Sigma^{-1} \underline{\mu}_2 + \underline{\mu}_2' \Sigma^{-1} \underline{\mu}_2) \\ \ln\left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right) &= -\frac{1}{2}(\underline{x}' \Sigma^{-1} \underline{x} - 2\underline{x}' \Sigma^{-1} \underline{\mu}_1 + \underline{\mu}_1' \Sigma^{-1} \underline{\mu}_1) + \frac{1}{2}(\underline{x}' \Sigma^{-1} \underline{x} - 2\underline{x}' \Sigma^{-1} \underline{\mu}_2 + \underline{\mu}_2' \Sigma^{-1} \underline{\mu}_2) \\ \ln\left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right) &= -\frac{1}{2}\underline{x}' \Sigma^{-1} \underline{x} + \underline{x}' \Sigma^{-1} \underline{\mu}_1 - \frac{1}{2}\underline{\mu}_1' \Sigma^{-1} \underline{\mu}_1 + \frac{1}{2}\underline{x}' \Sigma^{-1} \underline{x} - \underline{x}' \Sigma^{-1} \underline{\mu}_2 + \frac{1}{2}\underline{\mu}_2' \Sigma^{-1} \underline{\mu}_2 \\ \ln\left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right) &= \underline{x}' \Sigma^{-1} \underline{\mu}_1 - \underline{x}' \Sigma^{-1} \underline{\mu}_2 - \frac{1}{2}[\underline{\mu}_1' \Sigma^{-1} \underline{\mu}_1 - \underline{\mu}_2' \Sigma^{-1} \underline{\mu}_2] \\ \ln\left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right) &= \underline{x}' \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2) - \frac{1}{2}[\underline{\mu}_1' \Sigma^{-1} \underline{\mu}_1 - \underline{\mu}_2' \Sigma^{-1} \underline{\mu}_2] \end{aligned}$$

Novamente, como Σ^{-1} é matriz simétrica então:

$$\ln\left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right) = (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} \underline{x} - \frac{1}{2} [\underline{\mu}_1' \Sigma^{-1} \underline{\mu}_1 - \underline{\mu}_2' \Sigma^{-1} \underline{\mu}_2]$$

Completando-se o quadrado dos termos entre colchetes, tem-se:

$$\ln\left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right) = (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} \underline{x} - \frac{1}{2} [\underline{\mu}_1' \Sigma^{-1} \underline{\mu}_1 - \underline{\mu}_1' \Sigma^{-1} \underline{\mu}_2 + \underline{\mu}_1' \Sigma^{-1} \underline{\mu}_2 - \underline{\mu}_2' \Sigma^{-1} \underline{\mu}_2]$$

Lembrando, Σ^{-1} é matriz simétrica.

$$\ln\left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right) = (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} \underline{x} - \frac{1}{2} [\underline{\mu}_1' \Sigma^{-1} \underline{\mu}_1 - \underline{\mu}_1' \Sigma^{-1} \underline{\mu}_2 + \underline{\mu}_2' \Sigma^{-1} \underline{\mu}_1 - \underline{\mu}_2' \Sigma^{-1} \underline{\mu}_2]$$

$$\ln\left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right) = (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} \underline{x} - \frac{1}{2} [\underline{\mu}_1' \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2) + \underline{\mu}_2' \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2)]$$

$$\ln\left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right) = (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} \underline{x} - \frac{1}{2} [(\underline{\mu}_1' \Sigma^{-1} + \underline{\mu}_2' \Sigma^{-1})(\underline{\mu}_1 - \underline{\mu}_2)]$$

$$\ln\left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right) = (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} \underline{x} - \frac{1}{2} (\underline{\mu}_1' + \underline{\mu}_2') \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2)$$

$$\ln\left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right) = (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} \underline{x} - \frac{1}{2} (\underline{\mu}_1 + \underline{\mu}_2)' \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2)$$

Ou

$$\ln\left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right) = (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} \underline{x} - \frac{1}{2} (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} (\underline{\mu}_1 + \underline{\mu}_2)$$

Assim, se

$$(\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} \underline{x} - \frac{1}{2} (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} (\underline{\mu}_1 + \underline{\mu}_2) \geq \ln\left[\left(\frac{c(1/2)}{c(2/1)}\right)\left(\frac{p_2}{p_1}\right)\right]$$

então, \underline{x} pertence à região R_1 caso contrário pertence à região R_2 .

2.4 FUNÇÃO DISCRIMINANTE QUADRÁTICA

A função discriminante quadrática caracteriza-se por não fazer restrições à igualdade entre matrizes de covariância, embora mostre uma vulnerabilidade maior do que a análise discriminante linear à violação da hipótese de não normalidade, razão pela qual ela raramente é utilizada em estudos aplicados.

2.4.1 Classificação quadrática: populações com covariâncias diferentes: $\Sigma_1 \neq \Sigma_2$

Considere $f_1(\underline{x})$ e $f_2(\underline{x})$ funções normais multivariadas:

- A primeira com vetor de médias $\underline{\mu}_1$ e matriz de covariância Σ_1 ;
- A segunda com vetor de médias $\underline{\mu}_2$ e matriz de covariância Σ_2 e que $\Sigma_1 \neq \Sigma_2$.

A regra de classificação com base em ECM é a seguinte:

Resultado 3:

Se:

$$-\frac{1}{2} \ln \frac{|\Sigma_1|}{|\Sigma_2|} - \frac{1}{2} (\underline{\mu}_1' \Sigma_1^{-1} \underline{\mu}_1 - \underline{\mu}_2' \Sigma_2^{-1} \underline{\mu}_2) - \frac{1}{2} \underline{x}' (\Sigma_1^{-1} - \Sigma_2^{-1}) \underline{x} + (\underline{\mu}_1' \Sigma_1^{-1} - \underline{\mu}_2' \Sigma_2^{-1}) \underline{x} - \ln \left[\left(\frac{c(1/2)}{c(2/1)} \right) \left(\frac{p_2}{p_1} \right) \right] \geq 0$$

então \underline{x} pertence à região R_1 caso contrário pertence à região R_2 .

Neste caso, as regiões de classificação são definidas por funções quadráticas de \underline{x} .

Prova:

Sabe-se que as regiões R_1 e R_2 que minimizam o ECM são definidas pelos valores de \underline{x} tal que valem as desigualdades:

$$R_1: \frac{f_1(\underline{x})}{f_2(\underline{x})} \geq \left[\frac{c(1/2)}{c(2/1)} \right] \left[\frac{p_2}{p_1} \right] \quad \text{e} \quad R_2: \frac{f_1(\underline{x})}{f_2(\underline{x})} < \left[\frac{c(1/2)}{c(2/1)} \right] \left[\frac{p_2}{p_1} \right]$$

Onde:

$$\frac{f_1(\underline{x})}{f_2(\underline{x})} = \frac{\frac{1}{(2\pi)^{p/2} |\Sigma_1|^{1/2}} \exp\left[-\frac{1}{2}(\underline{x} - \underline{\mu}_1)' \Sigma_1^{-1} (\underline{x} - \underline{\mu}_1)\right]}{\frac{1}{(2\pi)^{p/2} |\Sigma_2|^{1/2}} \exp\left[-\frac{1}{2}(\underline{x} - \underline{\mu}_2)' \Sigma_2^{-1} (\underline{x} - \underline{\mu}_2)\right]}$$

Aplicando o logaritmo natural obtém-se:

$$\ln\left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right) \geq \ln\left[\left(\frac{c(1/2)}{c(2/1)}\right)\left(\frac{p_2}{p_1}\right)\right]$$

Desenvolvendo o primeiro membro da desigualdade temos:

$$\begin{aligned} \ln\left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right) &= \ln\left(\frac{1}{(2\pi)^{p/2} |\Sigma_1|^{1/2}} \exp\left[-\frac{1}{2}(\underline{x} - \underline{\mu}_1)' \Sigma_1^{-1} (\underline{x} - \underline{\mu}_1)\right]\right) \\ &\quad - \ln\left(\frac{1}{(2\pi)^{p/2} |\Sigma_2|^{1/2}} \exp\left[-\frac{1}{2}(\underline{x} - \underline{\mu}_2)' \Sigma_2^{-1} (\underline{x} - \underline{\mu}_2)\right]\right) \\ \ln\left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right) &= \ln\left(\frac{1}{(2\pi)^{p/2} |\Sigma_1|^{1/2}}\right) - \frac{1}{2}(\underline{x} - \underline{\mu}_1)' \Sigma_1^{-1} (\underline{x} - \underline{\mu}_1) \\ &\quad - \ln\left(\frac{1}{(2\pi)^{p/2} |\Sigma_2|^{1/2}}\right) + \frac{1}{2}(\underline{x} - \underline{\mu}_2)' \Sigma_2^{-1} (\underline{x} - \underline{\mu}_2) \\ \ln\left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right) &= \ln\left(\frac{1}{(2\pi)^{p/2}}\right) - \ln(|\Sigma_1|^{1/2}) - \frac{1}{2}(\underline{x} - \underline{\mu}_1)' \Sigma_1^{-1} (\underline{x} - \underline{\mu}_1) \end{aligned}$$

$$\begin{aligned}
\ln\left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right) &= \ln\left(\frac{1}{(2\pi)^{p/2}}\right) - \ln\left(|\Sigma_1|^{1/2}\right) - \frac{1}{2}(\underline{x} - \underline{\mu}_1)' \Sigma_1^{-1} (\underline{x} - \underline{\mu}_1) \\
&\quad - \ln\left(\frac{1}{(2\pi)^{p/2}}\right) + \ln\left(|\Sigma_2|^{1/2}\right) + \frac{1}{2}(\underline{x} - \underline{\mu}_2)' \Sigma_2^{-1} (\underline{x} - \underline{\mu}_2) \\
\ln\left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right) &= -\ln\left(|\Sigma_1|^{1/2}\right) - \frac{1}{2}(\underline{x} - \underline{\mu}_1)' \Sigma_1^{-1} (\underline{x} - \underline{\mu}_1) + \ln\left(|\Sigma_2|^{1/2}\right) + \frac{1}{2}(\underline{x} - \underline{\mu}_2)' \Sigma_2^{-1} (\underline{x} - \underline{\mu}_2) \\
\ln\left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right) &= -\ln\left(|\Sigma_1|^{1/2}\right) + \ln\left(|\Sigma_2|^{1/2}\right) - \frac{1}{2}(\underline{x} - \underline{\mu}_1)' \Sigma_1^{-1} (\underline{x} - \underline{\mu}_1) + \frac{1}{2}(\underline{x} - \underline{\mu}_2)' \Sigma_2^{-1} (\underline{x} - \underline{\mu}_2) \\
\ln\left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right) &= -\left(\ln\left(|\Sigma_1|^{1/2}\right) - \ln\left(|\Sigma_2|^{1/2}\right)\right) - \frac{1}{2}(\underline{x} - \underline{\mu}_1)' \Sigma_1^{-1} (\underline{x} - \underline{\mu}_1) + \frac{1}{2}(\underline{x} - \underline{\mu}_2)' \Sigma_2^{-1} (\underline{x} - \underline{\mu}_2) \\
\ln\left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right) &= -\left(\ln\frac{|\Sigma_1|^{1/2}}{|\Sigma_2|^{1/2}}\right) - \frac{1}{2}(\underline{x} - \underline{\mu}_1)' \Sigma_1^{-1} (\underline{x} - \underline{\mu}_1) + \frac{1}{2}(\underline{x} - \underline{\mu}_2)' \Sigma_2^{-1} (\underline{x} - \underline{\mu}_2) \\
\ln\left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right) &= -\frac{1}{2} \ln\frac{|\Sigma_1|}{|\Sigma_2|} + \left(-\frac{1}{2}(\underline{x} - \underline{\mu}_1)' \Sigma_1^{-1} (\underline{x} - \underline{\mu}_1) + \frac{1}{2}(\underline{x} - \underline{\mu}_2)' \Sigma_2^{-1} (\underline{x} - \underline{\mu}_2)\right) \\
\ln\left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right) &= -\frac{1}{2} \ln\frac{|\Sigma_1|}{|\Sigma_2|} - \frac{1}{2}\left((\underline{x} - \underline{\mu}_1)' \Sigma_1^{-1} (\underline{x} - \underline{\mu}_1) - (\underline{x} - \underline{\mu}_2)' \Sigma_2^{-1} (\underline{x} - \underline{\mu}_2)\right) \\
\ln\left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right) &= -\frac{1}{2} \ln\frac{|\Sigma_1|}{|\Sigma_2|} - \frac{1}{2}\left(\underline{x}' \Sigma_1^{-1} \underline{x} - \underline{x}' \Sigma_1^{-1} \underline{\mu}_1 - \underline{\mu}_1' \Sigma_1^{-1} \underline{x} + \underline{\mu}_1' \Sigma_1^{-1} \underline{\mu}_1\right) \\
&\quad + \frac{1}{2}\left(\underline{x}' \Sigma_2^{-1} \underline{x} - \underline{x}' \Sigma_2^{-1} \underline{\mu}_2 - \underline{\mu}_2' \Sigma_2^{-1} \underline{x} + \underline{\mu}_2' \Sigma_2^{-1} \underline{\mu}_2\right)
\end{aligned}$$

Como Σ_1^{-1} e Σ_2^{-1} são matrizes simétricas então:

$$\begin{aligned}
\ln\left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right) &= -\frac{1}{2} \ln\frac{|\Sigma_1|}{|\Sigma_2|} - \frac{1}{2}\left(\underline{x}' \Sigma_1^{-1} \underline{x} - \underline{x}' \Sigma_1^{-1} \underline{\mu}_1 - \underline{x}' \Sigma_1^{-1} \underline{\mu}_1 + \underline{\mu}_1' \Sigma_1^{-1} \underline{\mu}_1\right) \\
&\quad + \frac{1}{2}\left(\underline{x}' \Sigma_2^{-1} \underline{x} - \underline{x}' \Sigma_2^{-1} \underline{\mu}_2 - \underline{x}' \Sigma_2^{-1} \underline{\mu}_2 + \underline{\mu}_2' \Sigma_2^{-1} \underline{\mu}_2\right)
\end{aligned}$$

$$\begin{aligned}
\ln\left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right) &= -\frac{1}{2} \ln \frac{|\Sigma_1|}{|\Sigma_2|} - \frac{1}{2} \left(\underline{x}' \Sigma_1^{-1} \underline{x} - 2 \underline{x}' \Sigma_1^{-1} \underline{\mu}_1 + \underline{\mu}_1' \Sigma_1^{-1} \underline{\mu}_1 \right) \\
&+ \frac{1}{2} \left(\underline{x}' \Sigma_2^{-1} \underline{x} - 2 \underline{x}' \Sigma_2^{-1} \underline{\mu}_2 + \underline{\mu}_2' \Sigma_2^{-1} \underline{\mu}_2 \right) \\
\ln\left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right) &= -\frac{1}{2} \ln \frac{|\Sigma_1|}{|\Sigma_2|} - \frac{1}{2} \underline{x}' \Sigma_1^{-1} \underline{x} + \underline{x}' \Sigma_1^{-1} \underline{\mu}_1 - \frac{1}{2} \underline{\mu}_1' \Sigma_1^{-1} \underline{\mu}_1 + \frac{1}{2} \underline{x}' \Sigma_2^{-1} \underline{x} - \underline{x}' \Sigma_2^{-1} \underline{\mu}_2 \\
&+ \frac{1}{2} \underline{\mu}_2' \Sigma_2^{-1} \underline{\mu}_2 \\
\ln\left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right) &= -\frac{1}{2} \ln \frac{|\Sigma_1|}{|\Sigma_2|} - \frac{1}{2} \underline{x}' \Sigma_1^{-1} \underline{x} + \frac{1}{2} \underline{x}' \Sigma_2^{-1} \underline{x} + \underline{x}' \Sigma_1^{-1} \underline{\mu}_1 - \underline{x}' \Sigma_2^{-1} \underline{\mu}_2 - \frac{1}{2} \underline{\mu}_1' \Sigma_1^{-1} \underline{\mu}_1 \\
&+ \frac{1}{2} \underline{\mu}_2' \Sigma_2^{-1} \underline{\mu}_2
\end{aligned}$$

Novamente, como Σ_1^{-1} e Σ_2^{-1} são matrizes simétricas então:

$$\begin{aligned}
\ln\left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right) &= -\frac{1}{2} \ln \frac{|\Sigma_1|}{|\Sigma_2|} - \frac{1}{2} \left[\underline{x}' \Sigma_1^{-1} \underline{x} - \underline{x}' \Sigma_2^{-1} \underline{x} \right] + \underline{\mu}_1' \Sigma_1^{-1} \underline{x} - \underline{\mu}_2' \Sigma_2^{-1} \underline{x} \\
&- \frac{1}{2} \left[\underline{\mu}_1' \Sigma_1^{-1} \underline{\mu}_1 - \underline{\mu}_2' \Sigma_2^{-1} \underline{\mu}_2 \right] \\
\ln\left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right) &= -\frac{1}{2} \ln \frac{|\Sigma_1|}{|\Sigma_2|} - \frac{1}{2} \underline{x}' (\Sigma_1^{-1} - \Sigma_2^{-1}) \underline{x} + (\underline{\mu}_1' \Sigma_1^{-1} - \underline{\mu}_2' \Sigma_2^{-1}) \underline{x} - \frac{1}{2} (\underline{\mu}_1' \Sigma_1^{-1} \underline{\mu}_1 - \underline{\mu}_2' \Sigma_2^{-1} \underline{\mu}_2)
\end{aligned}$$

Assim, se:

$$\begin{aligned}
& -\frac{1}{2} \ln \frac{|\Sigma_1|}{|\Sigma_2|} - \frac{1}{2} (\underline{\mu}_1' \Sigma_1^{-1} \underline{\mu}_1 - \underline{\mu}_2' \Sigma_2^{-1} \underline{\mu}_2) - \frac{1}{2} \underline{x}' (\Sigma_1^{-1} - \Sigma_2^{-1}) \underline{x} + (\underline{\mu}_1' \Sigma_1^{-1} - \underline{\mu}_2' \Sigma_2^{-1}) \underline{x} \\
& - \ln \left[\left(\frac{c(1/2)}{c(2/1)} \right) \left(\frac{p_2}{p_1} \right) \right] \geq 0
\end{aligned}$$

então, \underline{x} pertence à região R_1 caso contrário pertence à região R_2 .

A classificação com funções quadráticas é bastante complicada e pode levar a diversos resultados estranhos. Isto é particularmente correto quando os dados não são

(essencialmente) normais.

Segundo JOHNSON e WICHERN [1998], se os dados não são normais, existem duas opções:

- Primeira, os dados podem ser transformados para dados quase normais, e um teste de igualdade de matrizes de covariância pode ser conduzido para ver se a regra linear ou a regra quadrática é apropriada. As transformações mais utilizadas foram propostas por *BOX-COX*¹. Porém essas transformações são possíveis em situações em que a variável dependente é conhecida e positiva.
- Segunda, pode-se usar uma função linear (ou quadrática) sem se preocupar sobre a forma da distribuição das populações e esperar que o procedimento funcione razoavelmente bem. Estudos realizados mostraram, porém, que existem casos de não normalidade onde a função de reconhecimento e classificação linear apresentou resultados ruins, embora as de matrizes de covariâncias fossem as mesmas. O ideal é sempre verificar o desempenho de um procedimento de classificação.

Segundo HAIR *et. al.* [2005], existindo dados suficientes pode-se separar os mesmos em duas amostras, uma de “treinamento” e outra de “teste”. As amostras de treinamento podem ser usadas para desenvolver a função de reconhecimento e classificação, e as amostras de validação podem ser usadas para avaliação de seu desempenho.

2.5 FUNÇÃO DISCRIMINANTE LINEAR DE FISHER

A Função Discriminante Linear de Fisher foi o primeiro método estatístico de discriminação e classificação. A idéia principal foi transformar a observação multivariada \underline{X} , de dimensão p , na observação univariada y (score), tal que os escores obtidos para as populações em estudo sejam separados tanto quanto possível [GUIMARÃES e CHAVES NETO, 2002].

¹Transformações que podem ser usadas em qualquer conjunto de dados onde a variável dependente seja positiva. O método consiste na aplicação de uma transformação exponencial nos dados de forma a aproximá-lo de uma situação normal. Foi desenvolvido por Box e Cox em 1964 [CARREIRA, 2005].

Assim, a Função Discriminante Linear de Fisher consiste em separar duas ou mais classes de indivíduos (pessoas, plantas, coisas ou tudo o que for objeto de estudo) ou fixar um novo objeto em uma das classes.

Para GUIMARÃES [2000], uma característica desta função é apresentar boas propriedades para a discriminação entre populações com a mesma matriz de covariância, pois a idéia básica é a de criar uma combinação linear das variáveis independentes de tal forma a definir a variável dependente.

Segundo LAM [*apud* SANTOS e MILIONI, 2006], a Função Discriminante Linear de Fisher se esforça em prover uma função linear pela qual se associam valores a dois ou mais atributos independentes, os quais são combinados produzindo uma simples pontuação de classificação. No caso de separação e classificação entre dois grupos, esta pontuação se compara a um valor de corte que separa os mesmos, permitindo então estabelecer a relação de pertinência do indivíduo a um dos grupos. Temos, portanto, uma equação linear do tipo $L = b_1x_1 + b_2x_2 + \dots + b_nx_n + c$ onde os coeficientes b_i são calculados de forma a maximizar a razão entre a variância entre os grupos e a variância entre os indivíduos do grupo e c é uma constante semelhante ao intercepto de uma regressão linear.

2.5.1 Função discriminante linear para duas populações

Neste item é apresentado a Função Discriminante Linear para Duas Populações descrita por GUIMARÃES e CHAVES NETO [2002]

Uma Função Discriminante Linear (FDL) tem a forma:

$$Z = \beta_0 + \beta_1X_1 + \dots + \beta_iX_i + \dots + \beta_{p-1}X_{p-1} = \beta_0 + \sum_{i=1}^{p-1} \beta_iX_i \quad (2.1)$$

onde Z é o escore discriminante (variável dependente) e os valores $\beta_i, i = 0, 1, \dots, p$ são os coeficientes da função nas variáveis $X_i, i = 0, 1, 2, \dots, p$, com $X_0 = 1$. A função retorna um valor Z , para uma nova observação \underline{m}_i , que é um vetor com as

medidas. O valor de Z funcionará como escore de classificação desta nova observação.

A Função Discriminante Linear transforma a observação multivariada \underline{X} , de dimensão p , na observação univariada y (escore), tal que os escores obtidos para as populações W_1 e W_2 sejam separados ao máximo. Sendo $\underline{\mu}_1$, $\underline{\mu}_2$ e Σ respectivamente os vetores médios de W_1 e de W_2 e a matriz de covariância comum a ambas as populações, tem-se a função 2.2:

$$y = (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} \underline{X} \quad (2.2)$$

Então, pode-se expressar a regra de classificação para \underline{X}_0 como:

Alocar \underline{X}_0 em W_1 se $y_0 - m \geq 0$ ou alocar \underline{X}_0 em W_2 se $y_0 - m < 0$

onde m é o valor de corte. Na realidade, os parâmetros $\underline{\mu}_1$, $\underline{\mu}_2$ e Σ não são conhecidos. Assim trabalha-se com os seus estimadores: $\overline{\underline{X}}_1$, $\overline{\underline{X}}_2$ e S_p , obtidos de amostras aleatórias dos grupos W_1 e W_2 com tamanhos n_1 e n_2 , respectivamente. O estimador S_p , da matriz Σ , tem por expressão 2.3:

$$S_p = \frac{1}{n_1 + n_2 - 2} [(n_1 - 1)S_1 + (n_2 - 1)S_2] \quad (2.3)$$

onde S_1 e S_2 são as matrizes de covariâncias amostrais. Assim, obtém-se a chamada Função Discriminante Linear de Fisher Amostral cuja expressão é:

$$\hat{y} = (\overline{\underline{X}}_1 - \overline{\underline{X}}_2)' S_p^{-1} \underline{X} \quad (2.4)$$

O valor de corte m é estimado por:

$$\hat{m} = \frac{1}{2}(\bar{y}_1 + \bar{y}_2) \quad (2.5)$$

Onde $\bar{y}_1 + \bar{y}_2$ são as médias dos escores para W_1 e W_2 . A regra de classificação fica:

Alocar \underline{X}_0 em W_1 se $y_0 = (\bar{\underline{X}}_1 - \bar{\underline{X}}_2)' S_p^{-1} \underline{X}_0 \geq \hat{m}$

Alocar \underline{X}_0 em W_2 se $y_0 = (\bar{\underline{X}}_1 - \bar{\underline{X}}_2)' S_p^{-1} \underline{X}_0 < \hat{m}$

2.6 REGRESSÃO LOGÍSTICA

A técnica da regressão logística aplica-se à obtenção da probabilidade de que uma observação pertença a um determinado conjunto, em função do comportamento das variáveis independentes.

É importante destacar que, na regressão logística, não existem restrições com relação à normalidade multivariada na distribuição das variáveis independentes, nem com relação à igualdade das matrizes de covariâncias dos dois grupos, como ocorre na análise discriminante, na qual se pressupõe a existência destas condições.

O modelo Logístico é bastante utilizado em Biometria para modelar a probabilidade do desenvolvimento de uma doença, em um período de tempo definido. Sua função assume valores entre 0 e 1 e apresenta uma forma de S, inferindo-se que o risco aumenta rapidamente em certo intervalo de valores intermediários da variável independente e, então, permanece extremamente alto, em torno de 1, para valores altos dessa variável.

Em outras palavras, esse modelo é de grande aplicabilidade por considerar a natureza multivariada de uma questão na pesquisa epidemiológica. Assim, esse modelo é muito utilizado para descrever a probabilidade de ocorrência de um evento [COSTA, 1997].

Na presença de k variáveis independentes categóricas ou contínuas, o modelo Logístico estabelece uma relação entre essas k variáveis e a probabilidade de ocorrência de sucesso de uma variável dependente dicotômica.

Torna-se necessário estabelecer uma relação linear entre as variáveis independentes

e uma transformação da variável resposta. Segundo DEMÉTRIO [1993], as transformações mais conhecidas são as transformações logística, *probit* e complemento *log-log*, que são denominadas de funções de ligação.

Na utilização do modelo de regressão logística, pode-se estar interessado no efeito de um fator de risco específico ou na identificação de vários fatores associados com a variável resposta.

2.6.1 Regressão logística para resposta dicotômica

Antes de descrevermos a Regressão Logística é importante ressaltar que modelos de Regressão Linear não são apropriados para resolver problemas com variável resposta dicotômica, ou binária. No Modelo Linear Geral (MLG) a variável dependente é relacionada às variáveis independentes por uma expressão da forma de (2.6):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_{p-1} X_{p-1} + \varepsilon = \beta_0 + \sum_{i=1}^{p-1} \beta_i X_i + \varepsilon \quad (2.6)$$

A aplicação do MLG supõe que os resíduos ε , dados por (2.7),

$$\varepsilon = Y - \beta_0 - \sum_{i=1}^{p-1} \beta_i X_i \quad (2.7)$$

são normalmente distribuídos, com média 0 (zero) e variância constante. Entretanto, quando a variável é dicotômica, os resíduos são eles mesmos dicotômicos, assumindo apenas os valores 0, para classificação correta, ou 1, para classificação incorreta (ou vice-versa), ou seja, os resíduos para a variável resposta na forma binária seguem uma distribuição de Bernoulli, cuja variância não é constante. Além disto, é possível que o MLG retorne para a variável resposta um valor fora do intervalo $[0, 1]$.

O Modelo de Regressão Logística para resposta dicotômica é apresentado a seguir conforme descrito por GUIMARÃES e CHAVES NETO [2002].

A Regressão Logística (RL) busca um modelo que permita relacionar a variável resposta, Y , aos fatores X_1, \dots, X_{p-1} , que, supõe-se, influenciam as ocorrências do evento estudado. A variável resposta deve ser do tipo dicotômica, ou binária, assumindo apenas os valores 0 ou 1. Neste caso considera-se que existe interesse apenas na ocorrência, ou não, do evento em questão. No modelo logístico a relação é dada por uma função sigmóide, ou seja:

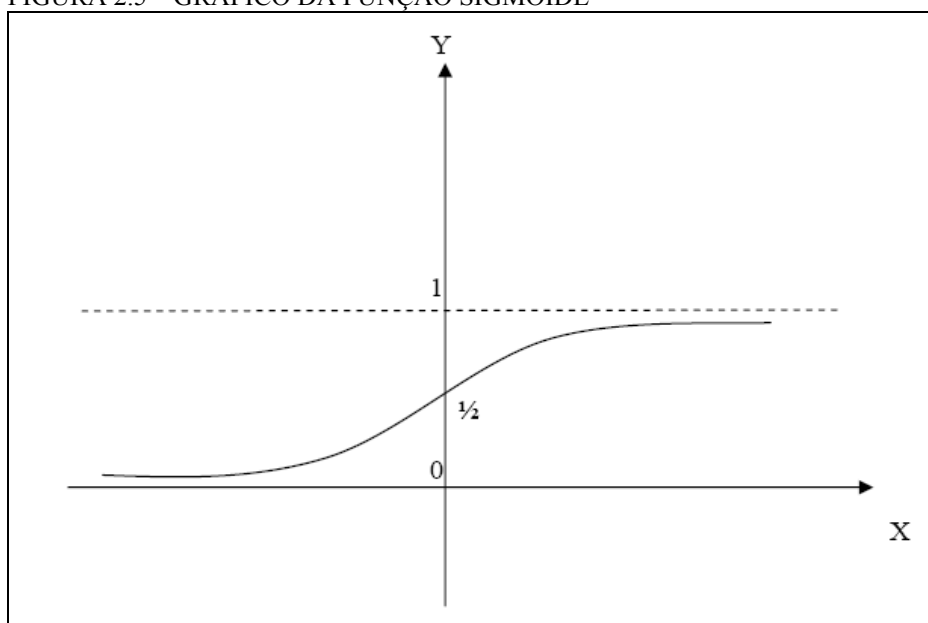
$$Y = \frac{e^{\mu}}{1 + e^{\mu}} \quad (2.8)$$

onde μ é uma expressão que relaciona os fatores acima e tem a forma:

$$\mu = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_{p-1} X_{p-1} + \varepsilon = \beta_0 + \sum_{i=1}^{p-1} \beta_i X_i + \varepsilon \quad (2.9)$$

A função (2.8) assume valores entre 0 e 1 para $\mu \in (-\infty, \infty)$, conforme a figura 2.5 a seguir.

FIGURA 2.5 – GRÁFICO DA FUNÇÃO SIGMÓIDE



FONTE: O AUTOR

A estimação dos coeficientes (parâmetros) $\beta_0, \beta_1, \dots, \beta_{p-1}$ é feita pelo Método da Máxima Verossimilhança, e dada a ocorrência da não-linearidade, o cálculo segue o Algoritmo de Levenberg–Marquardt..

2.6.2 Modelo de regressão logístico simples

O modelo de Regressão Logística Simples é descrito por LIMA [2002]:

Seja a amostra aleatória composta de n pares de observações (X_i, Y_i) com $i = 1, 2, \dots, n$, onde os Y 's representam os valores observados de uma variável dicotômica, e os X 's representam os valores observados de uma única variável independente, ou seja, $\mu = \beta_0 + \beta_1 X$. Assim a equação (2.6), quando Y é uma variável dicotômica, e tem-se apenas uma variável independente torna-se:

$$P(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}} \quad (2.10)$$

Esta expressão fornece a probabilidade condicional de que Y é igual a 1, dado o valor de X , ou seja, $P(Y = 1|X)$. Neste trabalho, isto significa dada uma certa característica (peso, por exemplo), a probabilidade da criança sobreviver no primeiro ano de vida.

Então a probabilidade condicional de que Y é igual a 0, fica

$$P(Y = 0|X) = 1 - P(X) \quad (2.11)$$

Portanto para as n observações têm-se:

$$P(Y_i = 1|X_i) = P(X_i) \text{ e } P(Y_i = 0|X_i) = 1 - P(X_i) \quad (2.12)$$

A abordagem do modelo de Regressão Logística baseia-se no método de Máxima Verossimilhança que fornece estimativas para os parâmetros que maximizam a probabilidade de obter o conjunto observado de dados.

Uma forma conveniente de expressar a contribuição da função de verossimilhança para os pares (X_i, Y_i) é através da fórmula (2.13):

$$F(X_i) = P(X_i)^{Y_i} [1 - P(X_i)]^{1-Y_i} \quad (2.13)$$

Sendo as observações independentes, a função de verossimilhança é obtida por (2.14):

$$l(\beta) = \prod_{i=1}^n F(X_i) = \prod_{i=1}^n [P(X_i)]^{Y_i} [1 - P(X_i)]^{1-Y_i} \quad (2.14)$$

Os estimadores de Máxima Verossimilhança destes parâmetros são escolhidos de forma a maximizar essa função, ou seja, deseja-se determinar o estimador para β que maximize a expressão (2.14). Matematicamente é mais fácil trabalhar com o logaritmo da mesma. A função obtida é chamada de função de log-verossimilhança e é dada por:

$$L(\beta) = \ln(l(\beta)) = \ln \left[\prod_{i=1}^n F(X_i) \right] = \ln \left[\prod_{i=1}^n [P(X_i)]^{Y_i} [1 - P(X_i)]^{1-Y_i} \right] \quad (2.15)$$

$$\text{Assim, } L(\beta) = \sum_{i=1}^n [Y_i \ln(P(X_i))] + (1 - Y_i) \ln[1 - P(X_i)] \quad (2.16)$$

$$\text{Então } L(\beta) = \sum_{i=1}^n \left[Y_i \ln \frac{e^{\beta'X}}{1 + e^{\beta'X}} + (1 - Y_i) \ln \left(1 - \frac{e^{\beta'X}}{1 + e^{\beta'X}} \right) \right] \quad (2.17)$$

$$\text{logo } L(\beta) = \sum_{i=1}^n \left[Y_i \ln \frac{e^{\beta'X}}{1 + e^{\beta'X}} + (1 - Y_i) \ln \left(\frac{1}{1 + e^{\beta'X}} \right) \right] \quad (2.18)$$

$$\text{Mas } \ln \left(\frac{e^{\beta'X}}{1 + e^{\beta'X}} \right) = \ln e^{\beta'X} - \ln(1 + e^{\beta'X}) = \beta'X - \ln(1 + e^{\beta'X}) \text{ e}$$

$$\ln \left(\frac{1}{1 + e^{\beta'X}} \right) = \ln 1 - \ln(1 + e^{\beta'X}) = -\ln(1 + e^{\beta'X})$$

que substituindo em (2.18) fica

$$L(\beta) = \sum_{i=1}^n [Y_i (\beta' X - \ln(1 + e^{\beta' X})) + (1 - Y_i)(-\ln(1 + e^{\beta' X}))] \quad (2.19)$$

Assim

$$L(\beta) = \sum_{i=1}^n [Y_i \beta' X - Y_i \ln(1 + e^{\beta' X}) - \ln(1 + e^{\beta' X}) + Y_i \ln(1 + e^{\beta' X})] \quad (2.20)$$

Portanto:

$$L(\beta) = \sum_{i=1}^n [Y_i \beta' X - \ln(1 + e^{\beta' X})] \quad (2.21)$$

Para determinar o valor de β que maximiza $L(\beta)$ deriva-se $L(\beta)$ em relação a β_0 e β_1 , igualando os resultados obtidos a zero. Obtem-se assim as equações de verossimilhança:

$$\frac{\partial L(\beta)}{\partial \beta_0} = \sum_{i=1}^n \left(Y_i - \frac{e^{\beta' X}}{1 + e^{\beta' X}} \right) = \sum_{i=1}^n [Y_i - P(X_i)] = 0 \quad (2.22)$$

e

$$\frac{\partial L(\beta)}{\partial \beta_1} = \sum_{i=1}^n \left(X_i \cdot Y_i - X_i \cdot \frac{e^{\beta' X}}{1 + e^{\beta' X}} \right) = \sum_{i=1}^n X_i \cdot [Y_i - P(X_i)] = 0 \quad (2.23)$$

Da equação (2.22) tem-se:

$$\sum_{i=1}^n Y_i = \sum_{i=1}^n \frac{e^{\beta' X}}{1 + e^{\beta' X}} = \sum_{i=1}^n P(X_i) \quad (2.24)$$

o que reflete no fato da soma dos valores observados de Y ser igual a soma dos valores esperados (preditos), mediante a aplicação do modelo obtido. Esta propriedade é importante, pois ajuda a avaliar o ajuste do modelo.

O valor de β dado pela solução das equações (2.22) e (2.23) é chamado estimador

de máxima verossimilhança e denotado por $\hat{\beta}'$. $P(X_i)$ representa uma estimativa da probabilidade condicional que Y é igual a 1, dado que X é igual a X_i .

2.6.3 Modelo de regressão logística múltiplo

Considerando-se uma amostra de n pares de observações independentes (X_i, Y_i) , $i = 1, 2, \dots, n$, da mesma forma como ocorreu no caso univariado, o ajuste do modelo requer que se obtenha uma estimativa do vetor $\underline{\beta}' = [\beta_0, \beta_1, \dots, \beta_p]$. A estimativa do vetor $\underline{\beta}'$ será feita, também, pelo método de Máxima Verossimilhança. A função de Verossimilhança é aproximadamente idêntica a dada na equação (2.10), fazendo-se a substituição de $\beta_0 + \beta_1 x$ por $\underline{\beta}' \underline{x}$, assim $P(X)$ é definido por:

$$P(X_i) = \frac{e^{\underline{\beta}' \underline{x}}}{1 + e^{\underline{\beta}' \underline{x}}} \quad (2.25)$$

As equações de verossimilhança, obtidas por derivadas da função de log-verossimilhança resultam em:

$$\sum_{i=1}^n [Y_i - P(X_i)] = 0 \quad (2.26)$$

e

$$\sum_{i=1}^n X_{ij} [Y_i - P(X_i)] = 0 \text{ para } j = 1, 2, \dots, p. \quad (2.27)$$

As soluções dessas equações de verossimilhança requerem a utilização de um método iterativo na busca dos parâmetros estimados $\hat{\beta}'s$.

Segundo GUIMARÃES [2000], considerando que o problema da estimação dos parâmetros pode ser entendido como um problema de minimização, pode-se utilizar algum método de minimização para a obtenção dos mesmos.

2.6.4 Métodos de estimação dos parâmetros por máxima verossimilhança (MV)

2.6.4.1 Método do gradiente

O método do gradiente é bastante simples em termos computacionais, mas tem convergência lenta chegando muitas vezes a não convergir em um tempo razoável; utiliza poucas informações, exige apenas as derivadas de primeira ordem para o cálculo do gradiente. Como o gradiente aponta na direção de maior crescimento da função no ponto, o método procura em cada ponto caminhar na direção oposta ao gradiente [LIMA, 2002]. Assim, a direção de busca é a direção oposta ao gradiente. Devido a sua simplicidade é bastante aplicável, porém, pode convergir muito lentamente, pois sua convergência é linear. O método do gradiente é definido pelo algoritmo iterativo:

$$X_{k+1} = X_k + \lambda_k d_k \quad (2.28)$$

Onde: λ_k é um escalar que minimiza $f(X + \lambda_k d_k)$, isto é, pesquisa-se a partir do ponto X_k ao longo da direção contrária ao gradiente da função até encontrar-se um mínimo nesta linha, onde este mínimo é X_{k+1} . Portanto este método propõe escolher a direção de descida (d_k), como a direção oposta ao Gradiente, ou seja, $d_k = -\nabla f(X_k)$.

2.6.4.2 Método de Newton

O método de Newton estende a idéia do método do gradiente aproveitando aproximações quadráticas da função f . Aproximações quadráticas não são somente melhores que aproximações lineares, mas ganham importância à medida que se aproximam do ponto ótimo do problema. Se f é uma função quadrática, o método de Newton determina o ponto de mínimo em um único passo. O valor de d_k tal que $\nabla^2 f(X_k).d_k = -\nabla f(X_k)$ é interpretado como uma correção na direção $-\nabla f(X_k)$ de forma a acelerar o processo [LIMA, 2002].

O método é executado em um processo iterativo tal que:

$$X_{k+1} = X_k + \lambda_k . d_k \quad (2.28)$$

Onde, as direções de busca são dadas por:

$$\nabla^2 f(X_k) . d_k = -\nabla f(X_k) \quad (2.29)$$

Apesar de ser um método de convergência quadrática local e rápida, deve-se considerar que para funções gerais, o método de Newton pode fornecer direções de pesquisa que não conduzam ao mínimo da função f .

2.6.4.3 Método de Levenberg-Marquardt

Este método trabalha com uma variação suave entre os extremos dos métodos de Newton e do Gradiente, sendo que utiliza este último quando está longe do mínimo, visto que o método de Newton pode não convergir nesta situação.

LIMA [2002] descreve que o método de Levenberg-Marquardt tem os seguintes procedimentos:

Os componentes da matriz Hessiana, mesmo não sendo utilizáveis de forma exata, têm alguma informação sobre a ordem de magnitude de escala para o problema. Considerando o método do Gradiente, onde a atualização é dada por: $X_{k+1} = X_k + \text{constante} . A$, pode-se fazer $\Delta X = X_{k+1} - X_k$, sendo um vetor adimensional que representa uma variação do vetor de parâmetros $\hat{\beta}'s$ do Modelo de Regressão Logístico Múltiplo a serem estimados.

O valor da função de log-verossimilhança é adimensional, isto é, é um número puro, bastando para isto ver a sua definição. Por outro lado, A_j tem a dimensão de $1/X_j$, o qual pode ser dimensional, isto é, tem unidades como un^{-1} (de fato, cada componente de A_j pode ter diferentes dimensões). A constante de proporcionalidade entre A_j e ΔX deve, portanto, ter a dimensão de X_j^2 . Na busca das unidades dos componentes da matriz Hessiana $[\alpha]$ pode-se ver que há apenas uma possibilidade para esta dimensão, ou seja, a unidade inversa do elemento da diagonal deve ser a escala da constante. Mas a escala pode, por si só, ser muito grande. Assim, divide-se a mesma por algum fator (adimensional)

λ , com a possibilidade do fator $\lambda \gg 1$ reduzir o passo. Em outras palavras a equação: $\Delta X_j = \text{constante } A_j$, pode ser trocada por:

$$\Delta X_j = \frac{1}{\lambda \cdot \alpha_{jj}} \cdot A_j \text{ ou } \lambda \cdot \alpha_{jj} \cdot \Delta X_j = A_j \quad (2.30)$$

Onde:

ΔX_j : Variação do parâmetro X_j ;

α_{jj} : Derivada segunda em relação ao parâmetro X_j .

A_j : Oposto da derivada primeira em relação ao parâmetro X_j , ou seja, $A_j = -\nabla f(X_j)$.

e necessariamente α_{jj} é positivo, o que é garantido pela equação:

$$\alpha_{lj} = \sum_{i=1}^n \left[\frac{\partial Y(X_i, \beta)}{\partial \beta_l} \cdot \frac{\partial Y(X_i, \beta)}{\partial \beta_j} \right] \text{ razão pela qual a mesma é adotada.}$$

As equações:

$$\lambda \cdot \alpha_{jj} \cdot \Delta_j \text{ e } \sum_{j=1}^p \alpha_{lj} \cdot \Delta X_j = A_l \quad (2.31)$$

que representam os métodos do gradiente e de Newton, podem ser combinadas ao definir-se uma nova matriz a com a seguinte forma

$$\alpha'_{jj} = \alpha_{jj} \cdot (1 + \lambda) \quad (2.32)$$

$$\alpha'_{lj} = \alpha_{lj} \text{ para } l \neq j \quad (2.33)$$

Assim, as equações anteriormente dadas (2.31) ficam:

$$\sum_{j=1}^p \alpha'_{ij} \Delta X_j = A_i \quad (2.34)$$

Quando λ é muito grande, a matriz Hessiana $[\alpha']$ é forçada a ser diagonal dominante e usa-se o método do gradiente. Por outro lado, quando λ aproxima-se de zero trabalha-se com o método de Newton.

2.7 AVALIAÇÃO DA EFICIÊNCIA

Um modo de avaliar a eficiência de modelos ou regras de reconhecimento de padrões e classificação é verificar a taxa de erro obtida na classificação dos indivíduos utilizando a função obtida (classificador). Essa taxa de erro será dada pela fração de observações erroneamente classificadas. Isso pode ser feito de duas formas.

A primeira, mais comum, consiste em separar o conjunto de dados em dois grupos. Um deles, chamado grupo de treinamento, é usado para desenvolver a função discriminante. O segundo grupo, chamado grupo de teste é usado para testar a função discriminante. Este método de validação da função discriminante é chamado de “Validação Cruzada” [HAIR, *et. al.*, 2005]. A taxa de erro obtida neste caso é a taxa de erro estimada.

Segundo GUIMARÃES e CHAVES NETO [2002], este procedimento apresenta dois inconvenientes. O primeiro está na redução do número de observações para a obtenção dos modelos, o que pode comprometer os resultados quando há poucas observações nos grupos estudados. O segundo inconveniente reside na possibilidade de formar a amostra de validação com as observações situadas na região de sobreposição das observações. Desta forma, ao proceder-se a avaliação, pode-se ter uma idéia equivocada da eficiência dos modelos.

A segunda maneira de avaliação da eficiência dos modelos consiste em usar todas as observações disponíveis para a construção da função de reconhecimento e classificação e em seguida classificar as mesmas observações disponíveis. Uma avaliação otimista irá ocorrer, pois os mesmos dados utilizados para a construção da função de reconhecimento e classificação serão usados para avaliá-la. A taxa de erro obtida, neste caso, é chamada de

Taxa de Erro Aparente – APER [JOHNSON e WICHERN, 1998].

Outro procedimento para avaliar modelos ou regras de reconhecimento e classificação de padrões é conhecido como Procedimento de Lachenbruch (1975) que consiste em, primeiramente, escolher um dos grupos e retirar uma observação do mesmo. Em seguida construir uma regra de classificação com as n_1-1 observações restantes do grupo escolhido e as n_2 observações do segundo grupo, obtendo uma função de reconhecimento e classificação. O próximo passo consiste em classificar a observação retirada, usando a função obtida anteriormente. Essa observação deve ser realocada no grupo de origem e repetem-se os procedimentos descritos para todas as outras observações dos dois grupos. Finalmente, constroem-se a regra de classificação com o total das $n = n_1 + n_2$ observações.

Desta forma, obtém-se uma regra de reconhecimento e classificação construída com as n observações amostrais e testada com todas as referidas observações. Este procedimento, contudo, pode ser inconveniente do ponto de vista computacional, em função do tempo de execução para a obtenção das n regras, uma a cada iteração.

2.8 A DISTRIBUIÇÃO NORMAL

Para o uso de técnicas de análise discriminante algumas premissas devem ser observadas. Segundo HAIR *et al.* [2005], os resultados da análise discriminante são afetados principalmente, no caso de violação das seguintes premissas:

1. Normalidade multivariada das variáveis independentes. Sua não observação pode causar problemas de estimação da função discriminante.
2. Matrizes de covariância iguais para os grupos definidos pelas variáveis dependentes. Amostras pequenas com matrizes de covariância desiguais prejudicam a classificação.

2.8.1 Distribuição normal multivariada

Conforme descreve MARQUES [2006], a generalização da distribuição normal

univariada para várias dimensões tem um papel fundamental na análise multivariada, pois grande parte das técnicas multivariadas aplicadas levam em consideração o fato da amostra possuir distribuição normal multivariada. Demonstra-se que a função densidade de uma normal multivariada é a generalização da função densidade da distribuição normal univariada para $p \geq 2$ dimensões.

Convenientemente denota-se a função densidade de probabilidade da distribuição normal, com média μ e variância σ^2 , por $X \sim N(\mu, \sigma^2)$. A distribuição normal univariada, com média μ e variância σ^2 , tem função densidade de probabilidade dada por (2.35).

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], x \in R, \mu \in R, \text{ e } \sigma \in R^+ \quad (2.35)$$

O desenvolvimento da função densidade de probabilidade conjunta da normal com p variáveis independentes normais X_1, X_2, \dots, X_p fica (2.36):

$$f(x_1, x_2, \dots, x_p) = \frac{1}{(2\pi)^{p/2} \sigma_1 \sigma_2 \dots \sigma_p} \exp\left[-\frac{1}{2} \sum_{i=1}^p \left(\frac{x_i - \mu_i}{\sigma_i}\right)^2\right] \quad (2.36)$$

JOHNSON e WICHERN [1998] demonstram que o termo:

$$\left(\frac{x - \mu}{\sigma}\right)^2 = (x - \mu)(\sigma^2)^{-1}(x - \mu) \quad (2.37)$$

no expoente da função densidade da normal, mede o quadrado da distância de x a μ em unidades de desvio padrão. Este termo pode ser generalizado para um vetor de observações $\underline{x}(p \times 1)$ de diversas variáveis, um vetor $\underline{\mu}(p \times 1)$ que representa o valor esperado do vetor aleatório \underline{X} , e uma matriz $\Sigma(p \times p)$ que é a matriz de variância-covariância de \underline{X} , conforme abaixo.

$$\underline{x}' = [x_1, x_2, \dots, x_p], \quad \underline{\mu}' = [\mu_1, \mu_2, \dots, \mu_p] \quad \text{e} \quad \Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_p^2 \end{bmatrix}, \quad \text{onde} \quad \sigma_{11} = \sigma_1^2,$$

$$\sigma_{22} = \sigma_2^2, \quad \sigma_{pp} = \sigma_p^2$$

Assim pode-se escrever (2.37) como:

$$(\underline{x} - \underline{\mu})' \Sigma^{-1} (\underline{x} - \underline{\mu}) \quad (2.38)$$

Agora pode-se escrever (2.36) como:

$$f(\underline{x}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp \left[-\frac{1}{2} (\underline{x} - \underline{\mu})' \Sigma^{-1} (\underline{x} - \underline{\mu}) \right] \quad (2.39)$$

onde: $-\infty < x_i < \infty, i = 1, 2, \dots, p$.

Assumindo que $\Sigma(p \times p)$ é qualquer matriz simétrica positiva definida (exemplificada em (2.40)), obtém-se a função densidade geral da normal multivariada descrita em (2.39). Denota-se a função densidade normal p-dimensional por $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$.

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_p^2 \end{bmatrix} \quad (2.40)$$

Feito o desenvolvimento da função densidade da normal multivariada $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$ com $|\Sigma| > 0$ prova-se o importante resultado a seguir, demonstrado em JOHNSON e WICHERN [1998]:

$$(\underline{x} - \underline{\mu})' \Sigma^{-1} (\underline{x} - \underline{\mu}) \sim \chi_p^2(\alpha) \quad (2.41)$$

com probabilidade $1 - \alpha$, que se pode denotar por:

$$P[(\underline{x} - \underline{\mu})' \Sigma^{-1} (\underline{x} - \underline{\mu}) \sim \chi_p^2(\alpha)] = 1 - \alpha \quad (2.42)$$

O resultado (2.42) tem importância fundamental, como metodologia, na avaliação da normalidade bivariada.

2.8.2 Avaliando a normalidade bivariada

Considerando-se $\underline{X} \sim N_2(\underline{\mu}, \Sigma)$ e substituindo no resultado (2.42) tem-se:

$$P[(\underline{x} - \underline{\mu})' \Sigma^{-1} (\underline{x} - \underline{\mu}) \sim \chi_2^2(0.5)] = 1 - \alpha = 1 - 0.5 = 0.5 \quad (2.43)$$

Com isso espera-se que 50% das observações amostrais situem-se dentro do contorno da elipse dada por:

$$(\underline{x} - \bar{\underline{x}})' S^{-1} (\underline{x} - \bar{\underline{x}}) \leq \chi_2^2(0.5) \quad (2.44)$$

onde se substitui $\underline{\mu}$ pelo seu estimador $\bar{\underline{x}}$ e Σ^{-1} pelo seu estimador S^{-1} , caso contrário a hipótese de normalidade é suspeita.

2.8.3 Avaliando a normalidade de uma distribuição com $p \geq 2$

Segundo JOHNSON e WICHERN [1998], um método formal utilizado para avaliar a normalidade de uma função com $p \geq 2$ é baseado no quadrado da distância generalizada, dado por:

$$d_j^2 = (\underline{x}_j - \bar{\underline{x}})' S^{-1} (\underline{x}_j - \bar{\underline{x}}), \text{ com } j = 1, 2, \dots, n \quad (2.45)$$

onde $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$ são as n observações amostrais.

Quando a população de onde a amostra foi retirada é normal multivariada sendo ambos n e $n - p$ maiores que 30 cada uma das distâncias quadradas $d_1^2, d_2^2, \dots, d_n^2$, comporta-se como uma variável aleatória qui-quadrado.

O método para avaliação da normalidade multivariada, neste caso, consiste nos seguintes passos:

1) Ordenar os quadrados das distâncias de forma crescente como

$$d_{(1)}^2 \leq d_{(2)}^2 \leq \dots \leq d_{(n)}^2;$$

2) Plotar os pares $\left[d_{(j)}^2, \chi_p^2 \left(\frac{j - \frac{1}{2}}{n} \right) \right]$.

Se o gráfico obtido resultar em uma faixa aproximadamente retilínea, assume-se a normalidade, caso contrário rejeita-se a normalidade.

2.9 MORTALIDADE INFANTIL: FATORES E INDICADORES

Entende-se por mortalidade infantil a terminologia utilizada para designar a morte de crianças em seu primeiro ano de vida e, também, o estudo estatístico referente ao número de crianças mortas numa determinada região no primeiro ano de vida.

A mortalidade infantil é um indicador de saúde importante e freqüentemente é usado por diversos pesquisadores da área. É calculado pela relação entre o número de óbitos de crianças menores de um ano e o número de nascidos vivos em um determinado local, sendo normalmente apresentado na base de mil nascidos vivos e é considerado um dos indicadores mais sensíveis da área de saúde. Conhecer seu perfil é fundamental para a formulação e adoção de estratégias que permitam o controle dos óbitos de menores de um ano. Essas estratégias vão desde uma assistência adequada à mulher durante a gravidez e o

parto até o acompanhamento pós-natal, principalmente, das crianças consideradas de risco.

CALDEIRA *et. al.* [2001], considera que a infância representa, de fato, uma fase particularmente vulnerável da vida, quando os determinantes biológicos do óbito estão fortemente atrelados às condições externas, sejam sócio-econômicas e ambientais (moradia, alimentação, saneamento, higiene e relações familiares), sejam referentes à disponibilidade de serviços de saúde.

Segundo TOMÉ e LATORRE [2006] é inegável que a mortalidade infantil apresenta relação com múltiplos fatores, tais como:

- Condições biológicas materna e infantil: idade da mãe, paridade, intervalo inter-partal, prematuridade, baixo peso ao nascer, retardo no crescimento intra-uterino, etc.;
- Condições ambientais: existência dos serviços de saúde e acessibilidade da população a eles, abastecimento de água potável e saneamento básico adequados, poluição, etc.; e
- Relações sociais que organizam a vida concreta das pessoas (moradia, trabalho, renda, nível de informação, proteção social, etc.).

CALDEIRA *et. al.* [2001], afirma que o Brasil possui níveis ainda elevados em mortalidade infantil, maiores inclusive do que os de alguns países da América Latina, apesar da queda acentuada ocorrida a partir da década de 90.

O desafio com relação à mortalidade infantil reside no aprimoramento da assistência pré-natal, parto e recém-nascido, principalmente prematuro.

2.9.1 A mortalidade infantil no mundo

As taxas de mortalidade infantil são muito elevadas nos países em desenvolvimento, já que são demonstrativas das suas condições deficientes de alimentação, de atendimento médico-hospitalar e de saneamento básico.

A mortalidade infantil é elevada em países em desenvolvimento e ocorre em função da pobreza e condições de subnutrição e de doenças infectocontagiosas a que estiverem sujeitos os bebês.

Nos últimos 50 anos, as taxas de mortalidade infantil no mundo em desenvolvimento foram reduzidas em 50% e as condições de saúde em todo o mundo aumentaram mais durante esse período do que em toda a história humana anterior. Entretanto, ANDRADE *et. al.* [2006] mostram que nas últimas décadas, mesmo em países de economia dependente ou onde persistem enormes desigualdades sociais ou, ainda, quando ocorrem períodos de crise econômica, tem-se observado decréscimo no valor dessa taxa, o que tem sido atribuído, principalmente, à redução da fecundidade e à expansão de serviços de saúde e de outros benefícios sociais, além de maior acesso a novas tecnologias em saúde.

No continente americano, também foi registrada notável redução da mortalidade infantil desde a segunda metade do século XX: de uma média de 90,34 óbitos de menores de um ano em cada mil nascidos vivos, nos anos 50, para uma taxa média de 31,31 nos anos 90 e, apesar dessa redução, as desigualdades no risco de morte infantil persistiram em todo o período, com uma mortalidade cerca de quatro a cinco vezes maior nos países mais pobres, em comparação aos mais ricos [ANDRADE *et. al.*, 2006].

Na América do Sul, o autor apresenta, segundo dados do Centro Latino-americano de Perinatologia e Desenvolvimento Humano (CLAP), a seguinte taxa de mortalidade infantil:

- Argentina (16,6 em 2000);
- Chile (10,1 em 1999);
- Paraguai (19,4 em 1999);
- Uruguai (14,1 em 2000) e;
- Venezuela (17,7 em 2000).

Estas taxas são, portanto, inferiores as apresentadas pelo Brasil em anos aproximados.

2.9.2 A mortalidade infantil no Brasil

O Brasil, assim como a maioria dos países, teve um decréscimo expressivo na taxa de mortalidade infantil de um modo geral, mas apresenta um grande contraste quando são comparadas as diversas regiões do país.

Segundo o JBONLINE [2004], a taxa nacional é considerada média dentro dos

parâmetros da Organização Mundial de Saúde (OMS). No Brasil, de 1991 a 1999, foi descrita uma redução de 28% no valor da taxa de mortalidade infantil, atingindo uma média de 31,8 por mil nascidos vivos no último ano analisado.

TABELA 2.1 – TAXA DE MORTALIDADE INFANTIL - BRASIL, GRANDES REGIÕES E ESTADOS:
1980 - 2004

BRASIL, GRANDES REGIÕES E UNIDADES DA FEDERAÇÃO	TAXA DE MORTALIDADE INFANTIL				RANKING EM 2004 ENTRE AS UNIDADES DA FEDERAÇÃO	PERCENTUAL DE QUEDA NO PERÍODO 1980 - 2004
	ANO DE REFERÊNCIA					
	1980	1991	2000	2004		
BRASIL	69.1	45.1	30.1	26.6		61.5
Norte	61.0	44.0	30.9	27.4	4º	55.1
Rondônia	55.5	38.5	29.1	25.9	13º	53.3
Acre	69.6	53.9	38.0	33.6	20º	51.7
Amazonas	58.2	42.5	32.3	28.5	16º	51.1
Roraima	70.8	40.9	22.9	20.7	6º	70.8
Pará	62.5	44.7	30.1	26.7	15º	57.3
Amapá	53.9	36.7	29.4	26.2	14º	51.4
Tocantins		49.1	33.6	29.9	17º	39.1
Nordeste	97.1	71.6	45.2	39.5	5º	59.3
Maranhão	86.1	73.6	49.9	43.6	26º	49.3
Piauí	81.0	61.9	36.5	31.6	18º	61.0
Ceará	111.5	71.1	38.1	33.2	19º	70.2
Rio Grande do Norte	111.2	72.1	44.7	38.8	23º	65.1
Paraíba	117.1	77.4	48.6	42.3	24º	63.9
Pernambuco	104.6	75.7	48.9	42.7	25º	59.2
Alagoas	111.6	98.5	63.8	55.7	27º	50.1
Sergipe	90.1	67.3	43.1	37.5	22º	58.4
Bahia	83.1	62.6	41.3	36.7	21º	55.9
Sudeste	57.7	31.8	22.2	19.5	2º	66.2
Minas Gerais	65.0	35.7	25.6	22.5	12º	65.4
Espírito Santo	48.2	31.7	23.5	20.7	8º	57.0
Rio de Janeiro	51.8	30.7	24.7	21.7	10º	58.1
São Paulo	56.7	29.9	19.4	17.0	2º	70.0
Sul	46.0	27.5	20.5	17.8	1º	61.3
Paraná	54.0	32.3	24.0	20.7	7º	61.7
Santa Catarina	46.1	28.6	20.3	17.7	3º	61.6
Rio Grande do Sul	36.5	21.6	16.7	14.7	1º	59.7
Centro - Oeste	47.9	32.5	23.3	20.7	3º	56.8
Mato Grosso do Sul	49.8	31.7	22.2	19.6	5º	60.7
Mato Grosso	49.2	35.3	25.0	22.2	11º	54.9
Goiás	47.3	33.2	23.9	21.3	9º	55.0
Distrito Federal	45.7	27.5	20.7	18.4	4º	59.8

FONTE: IBGE/2006

Segundo dados do IBGE [2006], em 2004 a taxa de mortalidade infantil do Brasil foi de 26,6 por mil habitantes, colocando o País na 99ª posição no ranking dos Países ou áreas com as mais baixas taxas estimadas pela ONU.

A Região Sul do Brasil, composta pelos Estados do Rio Grande do Sul, Santa

Catarina e Paraná, ainda que apresente o melhor desempenho em relação à mortalidade infantil no país, carece de estudos que avaliem diferenciais nos padrões dessa mortalidade segundo condições de vida.

De acordo com o MINISTÉRIO DA SAÚDE [2006], o Paraná é o Estado da Região Sul que apresenta valores mais elevados da Taxa de Mortalidade Infantil (19,6 por mil nascidos vivos em 2000), em comparação a 15,1 no Estado do Rio Grande do Sul e 15,9 em Santa Catarina.

Conhecer os diferenciais nos padrões dessa mortalidade, de acordo com as condições de vida e ambiente das comunidades, é fundamental para o estabelecimento de políticas públicas mais apropriadas a cada realidade. A tabela 2.1 apresenta os índices de mortalidade infantil nos Estados e nas Grandes Regiões, segundo o IBGE [2006].

2.9.2.1 Principais causas da mortalidade infantil

Os diversos estudos realizados refletem as condições de vida de uma população e se apóiam na influência que os determinantes do nível de vida exercem sobre a probabilidade de sobreviver ao primeiro ano de vida como destaca SOUZA e GOTLIEB [1993].

Segundo TOMÉ e LATORRE [2001], vários autores analisaram a tendência da mortalidade infantil associada a fatores de risco através de modelos de regressão e verifica-se que tanto os fatores relacionados à mãe, quanto os fatores relacionados à prestação de serviços à população foram os que mais apareceram em todos os estudos. Percebe-se que tais estudos visam alertar para as variáveis com alto grau na ocorrência de óbito, e assim, monitorar os recém-nascidos que devem receber assistência efetiva, contribuindo para a diminuição da mortalidade infantil.

O estudo realizado por KOZU *et. al.* [2006] mostra que, devido à grande quantidade de informações à respeito da mortalidade infantil, foi feita uma relação entre ela e as seguintes variáveis: história materna, doenças maternas, gravidez, pré-natal, fatores relacionados ao parto, fatores relacionados ao pós-parto e mortalidade infantil, descritos a seguir:

- História Materna: A história materna foi abordada em relação aos seguintes

aspectos: etnia, drogas lícitas e ilícitas, casamentos consangüíneos, estado marital e idade materna;

- Doenças Maternas: As principais doenças maternas associadas à mortalidade infantil encontradas no levantamento são: a hipertensão arterial, diabetes, infecção por HIV, sífilis e infecção por *streptococcus* grupo B;
- Gravidez: Foram analisados alguns estudos a respeito de fatores de risco para a mortalidade infantil relacionados à gravidez:
 - Pré-natal: Foi analisado que a ausência de cuidados pré-natais está associada com um aumento do risco de baixo peso ao nascer, partos prematuros e mortalidade materna e infantil;
 - Parto;
 - Pós-parto;
 - Mortalidade Perinatal.

MAGRO FILHO e RIBEIRO [2003] descrevem as principais causas de morte de menores de um ano e subdivide em três grupos:

- Crianças que morrem até 28 dias de vida: Neste primeiro grupo o autor subdivide em: 1) causas como referentes à má assistência ao pré-natal e ao parto em que se enquadra a falta de pré-natal adequado; má assistência ao parto como demora de atendimento, falta de vagas em hospitais; ausência de pessoal capacitado para o atendimento do recém-nascido; 2) Crianças com baixo peso (menos de 2.500g), nascidas de mães fumantes, alcoolistas ou drogadas; que tiveram graves problemas de saúde durante a gestação; gravemente desnutridas e; 3) Crianças com má formação congênita (proveniente de problemas gestacionais, algumas vezes produzidos por doenças infecciosas das mães); com má formação genética; e que nasceram prematuramente (antes de completarem 9 meses de gestação); nascidas de portadoras de HIV positivo (mães com AIDS), não tratadas;
- Crianças que morrem entre 28 dias e 1 ano de vida: estão neste grupo crianças que sobreviveram aos primeiros 27 dias de vida e morreram posteriormente em

consequência de doenças (patologias) anteriores; crianças que sofreram desmame precoce; crianças que não receberam as vacinas adequadas; crianças desnutridas (com maior facilidade de apresentar infecções); crianças com problemas respiratórios; crianças com doenças diarreicas (adquiridas por infecções transmitidas pela água, lixo ou falta de saneamento básico);

- Causas mal definidas, violência e acidentes: neste item o autor classifica as crianças que morrem em decorrência de maus tratos, acidentes domésticos e de trânsito, além de outras violências.

OLIVEIRA [2001] descreve algumas regras tanto para vivos quanto para óbitos no primeiro ano de vida.

Para vivos a autora descreve as seguintes regras:

- 1ª REGRA: se a mãe não teve filhos anteriormente, o peso da criança é alto e o Apgar no 5º minuto é alto, não há óbito;
- 2ª REGRA: se o Apgar no 5º minuto é alto e o peso da criança é alto, não ocorre óbito;
- 3ª REGRA: existe uma relação onde o Apgar no 1º minuto é alto, o peso é alto e não há óbito.

As regras que definem os óbitos no primeiro ano de vida são:

- 1ª REGRA: se a idade da mãe é de risco, o peso da criança é baixo, o Apgar no 1º minuto é baixo e o Apgar no 5º minuto é médio, ocorre óbito;
- 2ª REGRA: se o peso da criança é baixo e o Apgar no 1º e no 5º minuto são baixos, então há óbito;
- 3ª REGRA: se o Apgar no 1º minuto é baixo e o peso da criança é baixo, há óbito.

2.9.3 Índice de Apgar

O Índice de Apgar surgiu da necessidade de se ter um modo fácil e eficaz de avaliar o recém-nascido quanto as suas condições de saúde logo após seu nascimento. No ano de 1952, a Dra. Virginia Apgar apresentou e, posteriormente, publicou o seu sistema de avaliação do recém-nascido que é utilizado até os dias de hoje e desempenha um papel

fundamental no desenvolvimento das técnicas de atendimento ao recém-nascido em sala de parto [ABC DA SAÚDE, 2006].

Quando o bebê nasce, inicia-se a contagem do tempo e avalia-se o índice de Apgar no primeiro e no quinto minutos de vida da criança.

A Escala ou Índice de Apgar consiste na avaliação de 5 sinais objetivos do recém-nascido no primeiro e no quinto minuto após o nascimento, atribuindo-se a cada um dos sinais uma pontuação de 0 a 2, sendo utilizado para avaliar as condições dos recém-nascidos. Os sinais avaliados são: frequência cardíaca, respiração, tônus muscular, irritabilidade reflexa e cor da pele. O somatório da pontuação (no mínimo zero e no máximo dez) resultará no Índice de Apgar e o recém-nascido será classificado como sem asfixia (Apgar de 8 a 10), com asfixia leve (Apgar de 5 a 7), com asfixia moderada (Apgar de 3 a 4) e com asfixia grave (Apgar de 0 a 2).

No momento do nascimento, este índice é útil como parâmetro para avaliar as condições do recém-nascido e orientar nas medidas a serem tomadas, quando necessárias.

2.9.4 Coeficiente de mortalidade infantil

O coeficiente de mortalidade infantil é um indicador muito utilizado, e é tradicionalmente, dividido em mortalidade infantil neonatal e mortalidade infantil tardia que ocorrem, respectivamente, antes e após 28 dias. SANTA HELENA [2005], afirma que nas duas últimas décadas houve uma queda na mortalidade infantil, embora tenham existido períodos de estabilidade e até mesmo elevações dos coeficientes, sendo a mortalidade neonatal (óbito do nascido em até 27 dias completos de vida) se constitui no principal componente da mortalidade infantil.

No Brasil, nas últimas décadas, o coeficiente de mortalidade infantil, indicador utilizado internacionalmente para avaliar o nível de saúde de uma população, vem apresentando declínio, atribuído a uma concorrência de fatores como: intervenções ambientais, avanços da medicina, expansão no acesso e nos cuidados de saúde, diminuição da taxa de fecundidade, aumento do nível educacional da população, melhoria nutricional e das condições de vida, de uma maneira geral. Segundo estudiosos, em que pese a dissociação que

tem sido observada nas últimas décadas entre a redução da mortalidade infantil e melhorias sociais e econômicas, este coeficiente continua sendo utilizado, como marcador de desigualdade nas condições de vida das populações.

O Coeficiente de Mortalidade Infantil (CMI) representa a probabilidade de uma criança vir a falecer antes do seu primeiro aniversário. Classicamente, este coeficiente tem sido dividido em dois componentes: neonatal, que estima o risco de óbito nos primeiros 27 dias de vida, e o pós-neonatal, que mede o risco de óbito entre o 28º dia de vida e o final do primeiro ano de vida.

De acordo com NOVAES, *et. al.* [2002] o CMI também contribui para subsidiar processos de planejamento, gestão e avaliação de políticas e ações de saúde voltadas para a atenção pré-natal e ao parto, bem como para a proteção da saúde infantil.

As taxas de mortalidade infantil podem ser classificadas como:

- Altas: 50 ou mais por mil;
- Médias: 20-49 por mil;
- Baixas: menos de 20 por mil.

Esses parâmetros devem ser periodicamente ajustados às mudanças verificadas no perfil epidemiológico em cada local; atualmente, vários países e poucos municípios brasileiros apresentam valores abaixo de 10 por mil. De maneira geral, quando a taxa de mortalidade infantil é alta, o componente pós-neonatal é predominante e quando a taxa é baixa, o seu principal componente é a mortalidade neonatal, com predomínio da mortalidade neonatal precoce. Altas taxas de mortalidade infantil refletem baixos níveis de saúde e de desenvolvimento sócio-econômico da população, o que pode também acontecer em alguns segmentos sociais, mesmo onde as taxas da população geral são reduzidas.

A taxa de mortalidade infantil pode ser calculada pelo método direto com a utilização do Sistema de Informações sobre Mortalidade (SIM) e Sistema de Informações sobre Nascidos Vivos (SINASC) do Ministério da Saúde, ou pelo método indireto (IBGE: estimativas baseadas no Censo Demográfico e na Pesquisa Nacional por Amostra de Domicílios - PNAD). O cálculo direto da taxa a partir de dados obtidos dos sistemas de informação do Ministério da Saúde pode exigir correções do sub-registro de óbitos infantis e

de nascidos vivos, especialmente nas regiões Norte e Nordeste.

2.9.5 Declaração de nascido vivo

O Sistema de Informações sobre Nascidos Vivos (SINASC) do Ministério da Saúde foi implantado no Brasil a partir de 1990, com o intuito de obter informações fidedignas sobre gestação, nascimentos e características maternas podendo-se, assim, através da elaboração de diversos indicadores, planejar ações de saúde destinadas à população materno-infantil [MINISTÉRIO DA SAÚDE, 2006].

Segundo MISHIMA [1999], o SINASC é um sistema de coleta de dados sobre nascimentos vivos, alternativo ao Registro Civil, utilizando como instrumento básico a DN, (ver ANEXO I – DECLARAÇÃO DE NASCIDO VIVO (DN)) documento padronizado e individualizado constituído de três vias, com cores e finalidades distintas:

- A primeira, branca, é enviada aos órgãos regionais das Secretarias Estaduais de Saúde, com finalidade estatística;
- A segunda, amarela, é destinada à família para apresentação ao Cartório, no momento do registro civil do nascimento, e;
- A terceira, rosa, também se destina à família para ser entregue ao serviço de saúde numa primeira consulta da criança, para fornecer informações sobre o seu nascimento.

Para essa terceira via, existe um fluxo alternativo, proposto pelo Ministério da Saúde, no qual o hospital a envia diretamente às unidades básicas de saúde, para efetiva vigilância do recém-nascido de alto risco.

A partir de sua implantação, a Declaração de Nascido Vivo (DN) tornou-se o instrumento obrigatório para a coleta de dados. O Sistema, implantado oficialmente a partir de 1990, foi desenvolvido à semelhança do Sistema de Mortalidade – SIM, com o objetivo de coletar dados sobre os nascimentos informados em todo território nacional e o fornecimento de dados sobre natalidade para todos os níveis do sistema de saúde. O documento de entrada do sistema é a Declaração de Nascido Vivo – DN, padronizada em todo o país.

O SINASC tem, entre seus objetivos, a utilização dos dados produzidos a nível

local, visando não só o conhecimento dos dados, mas também vigilância dos recém-nascidos de alto risco, através do monitoramento da realidade local e o planejamento de intervenções de acordo com as necessidades da população adstrita².

Segundo o MINISTÉRIO DA SAÚDE [2006], o embasamento legal para o funcionamento do sistema é decorrente dos seguintes documentos:

- 1) Lei nº 6.015, de 31 de dezembro de 1973, Lei dos Registros Públicos que, no seu Capítulo IV (artigos 51 a 67), define as normas e providências para registro dos nascimentos em todo o território nacional;
- 2) Decreto nº 4.726, de 09 de junho de 2003, do Sr. Presidente da República, que aprova a estrutura regimental do Ministério da Saúde – MS, criando a Secretaria de Vigilância em Saúde/SVS;
- 3) Portaria nº 20, de 03 de outubro de 2003, da Secretaria de Vigilância em Saúde - SVS regulamenta as rotinas de coleta de dados, fluxo e periodicidade de envio das informações sobre óbitos;
- 4) Portaria nº 1.929, de 09 de outubro de 2003, que define as atribuições da Secretaria de Vigilância em Saúde e do Departamento de Informática do SUS no que se refere aos Sistemas de Informações. Esta portaria, em seu artigo 1º, designa a SVS como órgão gestor no nível nacional do Sistema de Informações sobre Mortalidade. O Departamento de Análise da Situação de Saúde – DASIS, através da sua Coordenação Geral de Informações e Análise Epidemiológica – CGIAE, é o órgão da SVS responsável por este gerenciamento, aqui incluídas a definição de variáveis, críticas e agregações de dados, impressão e distribuição dos documentos de captação dos dados (declarações de nascido vivo), manuais de operação do sistema, bem como a ordenação de alterações que se fizerem necessárias;
- 5) Portaria nº 16, de 23 de abril de 2004, que constitui o Comitê Técnico Assessor do Sistema de Informações sobre Nascidos Vivos – CTA – SINASC e dá outras

² Dependente, estreitado, ligado, preso, submetido.

providências.

A implantação do SINASC ocorreu de forma lenta e gradual em todas as unidades da Federação e já vem apresentando em muitos municípios, desde o ano de 1994, um número maior de registros do que o publicado pelo IBGE com base nos dados de Cartório de Registro Civil, além de possibilitar a construção de indicadores úteis para o planejamento e gestão dos serviços de saúde. A implantação do SINASC permitiu a obtenção de dados mais detalhados e fidedignos sobre nascidos vivos do que os existentes anteriormente. As instituições de saúde passaram a utilizar esses dados como fonte de pesquisas, avaliação epidemiológica e administrativa da assistência materno-infantil prestada em suas áreas.

CAPÍTULO III

3 MATERIAIS E MÉTODOS

As informações utilizadas para a realização do presente estudo são provenientes do banco de dados do SINASC e do SIM da 6ª Regional de Saúde de União da Vitória – PR referentes ao período de 1999 a 2005.

A 6ª Regional de Saúde caracteriza-se como um Consórcio Intermunicipal de Saúde, cuja finalidade é possibilitar a expansão e melhoria da qualidade dos serviços oferecidos à população, além de representar um processo social de articulação de interesses da região, tendo como ponto de partida a melhoria das condições de atendimento aos cidadãos.

A região que compreende a 6ª Regional de Saúde é composta por 09 municípios: União da Vitória, Paula Freitas, Paulo Frontin, General Carneiro, Cruz Machado, Porto Vitória, Bituruna, Antônio Olinto e São Mateus do Sul, abrangendo um total de 164.803 habitantes em uma área total de 7298 km².

3.1 DADOS DO ESTUDO

Os dados estudados³ constituíram-se de 18.770 nascimentos com Declaração de Nascido Vivo, ocorridos entre 1999 e 2005, filhos de mães residentes nos municípios que constituem a 6ª Regional de Saúde e de 226 óbitos infantis relacionados a essa coorte de nascidos vivos, com Declaração de Óbito (DO), ocorridos no mesmo período.

Na pesquisa realizada, muitas observações foram descartadas, pois estavam incompletas ou preenchidas com valores estranhos. Este fato prejudicou a construção da função discriminante, principalmente para o grupo de crianças com óbito no primeiro ano de vida, pois 22,60 % (66 em 292) das observações foram descartadas. No grupo de crianças com sobrevivência no primeiro ano de vida, o descarte realizado foi de 8,39% (1.698 em 20.242).

³ Disponível em: http://www.uniguacu.edu.br/prof_ivan

Foi realizado o procedimento de relacionamento entre os bancos de dados de óbitos e nascimentos a fim de identificar os nascidos vivos que evoluíram para óbito infantil, sendo utilizados os campos nome da mãe, data do nascimento e sexo para a verificação e emparelhamento da DN com sua respectiva DO (ver ANEXO II – DECLARAÇÃO DE ÓBITO (DO)).

Criaram-se através deste procedimento duas bases de dados para o estudo: os nascidos vivos que ultrapassaram o primeiro ano de vida e os que entraram em óbito neste período.

3.2 VARIÁVEIS ANALISADAS

A variável dependente do estudo foi a ocorrência de morte ou não antes de completar o primeiro ano de vida e as variáveis independentes ou explicativas foram:

- Sexo: 1 para menino e 0 para menina;
- Peso ao nascer: medido em gramas;
- Índice de Apgar no primeiro e quinto minutos de vida: valores de 0 a 10;
- Tipo de gravidez: 1 para única, 2 para dupla, 3 para tripla ou mais;
- Duração da gestação: medida em uma escala de 1 a 6;
- Número de consultas pré-natal: medida numa escala de 1 a 4;
- Tipo de parto: 1 para vaginal e 2 para cesário;
- Grau de instrução da mãe: medida em uma escala de 1 a 5;
- Idade da mãe: medida em anos;
- Número de gestações anteriores: dado pelo número de filhos tidos em gestações anteriores, sobreviventes ou não;
- Estado civil: 1 para solteira, 2 para casada, 3 para viúva, 4 para separada e 5 para união consensual.

A utilização destes fatores foi considerada mais importante na determinação da probabilidade de óbito (ou não) do recém-nascido, pois favorece possibilidades na identificação de fatores determinantes da mortalidade infantil e auxilia na avaliação da qualidade de vida de uma população bem como a disponibilidade e a qualidade da assistência

prestada pelo serviço de saúde.

Segundo estudo realizado por SANTA HELENA *et. al.* [2005] os fatores socioeconômicos da mãe, como idade, estado civil e escolaridade, bem como os fatores de prematuridade, escore do Apgar, peso e tipo de gravidez, são fatores de risco descritos na literatura que apresentam maior associação com a mortalidade no período neonatal.

LUO e KALBERG [*apud* SANTA HELENA *et. al.*, 2005], demonstram que o baixo peso corporal e o Apgar baixo são apontados como os maiores preditores de mortalidade infantil, mas não devem ser apontados como fatores de risco isolados, e sim como mediadores que atuam em diversos determinantes e condicionantes do óbito infantil.

3.3 PROCESSAMENTO DOS DADOS

A etapa seguinte à obtenção dos dados foi a realização das análises estatísticas para a associação dos vários fatores preditores com a mortalidade infantil através da utilização da Função Discriminante Quadrática, não sendo utilizados os modelos convencionais utilizados na análise da mortalidade infantil, baseados em regressões logísticas, modelos *log*-lineares ou modelos de riscos.

O gerenciamento dos dados foi feito utilizando-se o programa de banco de dados *Access* 2003, sendo este o formato do banco de dados da 6ª Regional de Saúde. A seguir os dados foram exportados para planilhas do *Excel* e salvos em formato *.csv. Já na classificação das observações, foi utilizado o programa desenvolvido em Delphi 7.0, com suporte do banco de dados *Firebird* (Sistema Gerenciador de Banco de Dados versão *Open Source* do *Interbase* 6.0). O programa trabalhou com os dados salvos nos arquivos *.csv, através da associação das bases de dados de nascidos vivos e de óbitos, utilizando as variáveis registradas nas DN's e nas DO's, tendo a variável resposta binária (0 se a criança morreu e 1 se sobreviveu) (ver ANEXO III – CÓDIGO FONTE DO PROGRAMA).

O programa desenvolvido faz quatro importações de dados: duas importações referentes a observações de crianças sobreviventes e outras duas de crianças não sobreviventes. Cada grupo de duas observações refere-se a uma amostra de treinamento e uma amostra de testes. Em seguida, é calculado o vetor médio das observações e calculada a matriz

de covariância das matrizes de treinamento, que são as principais estatísticas do método. Utilizando-se o algoritmo de Gauss-Jordan são calculadas as matrizes inversas das matrizes de covariância. O passo seguinte consistiu na decomposição das matrizes de covariância utilizando-se o processo de decomposição de Cholesky, fornecendo outras duas matrizes que serão utilizadas para calcular o valor do determinante de cada matriz de covariância. Os resultados obtidos são utilizados para a construção da função discriminante quadrática. Cada observação das amostras de validação é inserida na função discriminante obtida. Seu resultado é avaliado. Se o valor obtido é positivo, a observação é classificada na população sobrevivente, caso contrário é classificado na população não sobrevivente (ver ANEXO IV - MANUAL DO PROGRAMA).

Para a identificação de eventuais erros foi feita conferência visual, assim como avaliação de erros de amplitude e de consistência lógica. Os dados também foram compilados com o software *MINITAB* para consistência dos resultados.

CAPÍTULO IV

4 RESULTADOS E DISCUSSÃO

4.1 ANÁLISES PRELIMINARES

A partir do programa desenvolvido para efetuar a construção da função de reconhecimento e classificação quadrática e avaliação de seu desempenho, realizou-se diversas análises com as amostras de treinamento e amostras de teste diferentes. O objetivo foi avaliar o desempenho da função de reconhecimento e classificação, calculando a Taxa Aparente de Erro e a Taxa de Erro Estimada. Não se usou o Procedimento de Lachenbruch, pois o tempo computacional seria extremamente alto (algo em torno de 130 horas para 18.770 regras construídas – Processador Athlon 64 3000 +). Também foram alterados os valores das probabilidades a priori. Foram tomadas probabilidades iguais a 50% (para as duas populações: sobrevivência e óbito no 1º ano de vida) e 1,2% para óbito e 98,8% para sobrevivência (no primeiro ano de vida).

Realizou-se, em seguida, novo processamento dos dados, excluindo-se uma a uma, as variáveis estudadas no decorrer do trabalho. Foram realizados outros dois processamentos, utilizando as variáveis importantes na determinação da mortalidade infantil, conforme descrito na literatura, e ainda, mais dois processamentos com nove e oito variáveis complementares ao conjunto total das variáveis estudadas em relação aos processamentos anteriores.

Os resultados foram usados para avaliar o desempenho da função de reconhecimento e classificação.

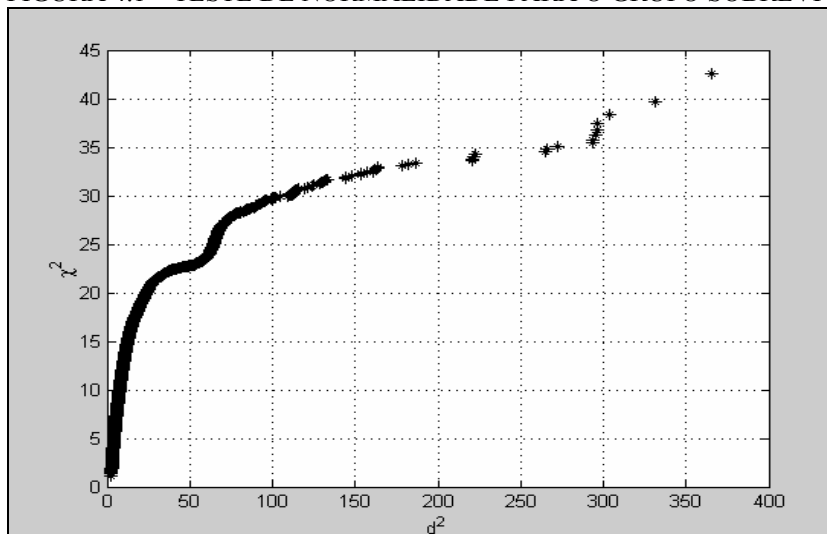
4.2 AVALIAÇÃO DA NORMALIDADE DOS DADOS

Utilizando o software *Matlab R12*, plotando o quadrado das diferenças com os respectivos valores de Qui-quadrado obteve-se os seguintes gráficos:

- 1) Observações do Grupo Sobrevivente ao Primeiro Ano de Vida, conforme figura

4.1:

FIGURA 4.1 – TESTE DE NORMALIDADE PARA O GRUPO SOBREVIVENTE

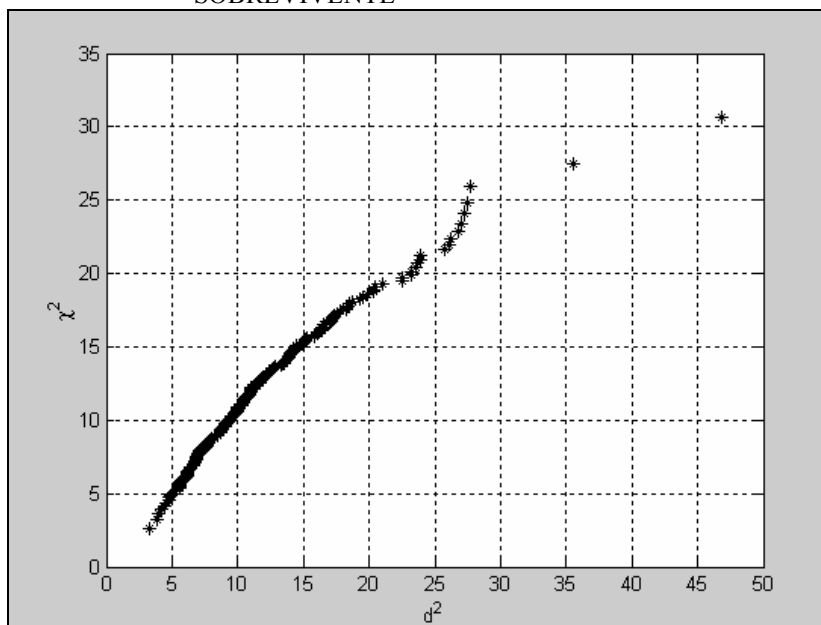


FONTE: O AUTOR

Observa-se que o gráfico sugere a não existência de normalidade dos dados.

2) Observações do Grupo Não-Sobrevivente ao Primeiro Ano de Vida, conforme figura 4.2:

FIGURA 4.2 – TESTE DE NORMALIDADE PARA O GRUPO NÃO-SOBREVIVENTE



FONTE: O AUTOR

Observa-se que o gráfico sugere, razoavelmente, a existência de normalidade dos dados.

4.3 ANÁLISE DO MODELO

Utilizou-se 5 amostras estratificadas para a construção da função de reconhecimento e treinamento com 60% dos dados e teste com 40% dos dados restantes, conforme sugere HAIR *et. al.* [2005]. Os resultados estão apresentados a seguir:

Resultados com probabilidades a priori iguais:

FIGURA 4.3 – RESULTADOS COM AMOSTRA 01 – PROBABILIDADE A PRIORI IGUAIS

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	25	5	2	5	1	2	4	0	0	2	3100	0			
1	34	2	4	5	1	2	4	0	8	10	3330	0			
1	18	2	4	5	1	1	4	0	9	10	3160	0			

Processamento

Realizar classificação

Probabilidade a priori de óbito: 0,5

Altera probabilidade a priori

Salvar classificação

Taxa aparente de erro: 8,73%

Importar: 4- Val. M

Importar

Limpar Dados

Log: 14:15:43 Calculando subtração das inversas das covariâncias
14:15:43 Obtendo resultados para vivos
14:15:43 --> Ln(priorim/prioriv): 0
14:15:44 Obtendo resultados para mortos
14:15:44 Fim da classificação

FONTE: O AUTOR

FIGURA 4.4 – RESULTADOS COM AMOSTRA 02 – PROBABILIDADE A PRIORI IGUAIS

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	25	1	3	5	1	1	4	1	8	9	3110	9			
4	16	1	4	2	1	1	2	0	1	8	600	1			
4	26	1	4	4	1	1	3	0	6	8	1700	5			

Processamento

Realizar classificação

Probabilidade a priori de óbito: 0,5

Altera probabilidade a priori

Salvar classificação

Taxa aparente de erro: 9,33%

Importar: 4- Val. M

Importar

Limpar Dados

Log: --> Progresso: 90
14:44:13 Calculando vetor médio
14:44:13 --> População vivos : 11126
14:44:14 --> População mortos : 136
14:44:14 Calculando covariância para vivos

FONTE: O AUTOR

FIGURA 4.5 – RESULTADOS COM AMOSTRA 03 – PROBABILIDADE A PRIORI IGUAIS

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	21	5	2	5	1	2	3	1	5	8	2470	4			
4	20	1	4	2	1	1	1	0	4	7	800	5			
4	20	1	4	5	1	2	3	0	9	10	3350	5			

Importar
4- Val. M
Importar
Limpar Dados

Processamento

Realizar classificação

Probabilidade a priori de óbito
0,5

Altera probabilidade a priori

Salvar classificação

Taxa aparente de erro 8,92%

☒ Covariância Vivos
☒ Covariância Mortos
☒ Inversão Vivos
☒ Inversão Mortos

☒ Cholesky Vivos
☒ Cholesky Mortos
☒ Determinantes
☒ Resultados

	Vivos	Óbitos
Dos Vivos	6770	638
Dos Mortos	31	59

15:00:59 Importando dados
--> Progresso: 90
15:01:06 Calculando vetor médio
15:01:06 --> População vivos : 11136
15:01:06 --> População mortos : 136

FONTE: O AUTOR

FIGURA 4.6 – RESULTADOS COM AMOSTRA 04 – PROBABILIDADE A PRIORI IGUAIS

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	28	1	3	5	1	1	4	1	7	8	4040	0			
4	37	1	3	4	1	1	3	0	7	8	2920	4			
4	25	5	2	5	1	2	4	0	0	2	3100	0			

Importar
4- Val. M
Importar
Limpar Dados

Processamento

Realizar classificação

Probabilidade a priori de óbito
0,5

Altera probabilidade a priori

Salvar classificação

Taxa aparente de erro 8,77%

☒ Covariância Vivos
☒ Covariância Mortos
☒ Inversão Vivos
☒ Inversão Mortos

☒ Cholesky Vivos
☒ Cholesky Mortos
☒ Determinantes
☒ Resultados

	Vivos	Óbitos
Dos Vivos	6793	625
Dos Mortos	34	56

15:13:39 Importando dados
--> Progresso: 90
15:13:45 Calculando vetor médio
15:13:46 --> População vivos : 11126
15:13:46 --> População mortos : 136

FONTE: O AUTOR

FIGURA 4.7 – RESULTADOS COM AMOSTRA 05 – PROBABILIDADE A PRIORI IGUAIS

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	34	5	3	2	1	1	2	0	4	1	850	2			
4	19	1	3	4	1	1	4	0	3	5	1950	0			
4	18	1	4	2	1	1	1	1	4	6	500	6			

Importar
4- Val. M
Importar
Limpar Dados

Processamento

Realizar classificação

Probabilidade a priori de óbito
0,5

Altera probabilidade a priori

☒ Covariação Vivos
☒ Covariação Mortos
☒ Inversão Vivos
☒ Inversão Mortos

☒ Cholesky Vivos
☒ Cholesky Mortos
☒ Determinantes
☒ Resultados

	Vivos	Óbitos
Dos Vivos	6809	609
Dos Mortos	33	57

Salvar classificação

Taxa aparente de erro 8,55%

15:24:37 Importando dados
--> Progresso: 90
15:24:45 Calculando vetor médio
15:24:45 --> População vivos : 11126
15:24:45 --> População mortos : 136

FONTE: O AUTOR

Considerando-se todos os dados disponíveis tanto para a construção da função de reconhecimento e classificação quanto para a validação da mesma, obtiveram-se os seguintes resultados:

FIGURA 4.8 – RESULTADOS COM UTILIZAÇÃO DE TODAS AS OBSERVAÇÕES – PROBABILIDADES A PRIORI IGUAIS

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	25	1	3	5	1	1	4	1	8	9	3110	9			
4	16	1	4	2	1	1	2	0	1	8	600	1			
4	26	1	4	4	1	1	3	0	6	8	1700	5			

Importar
4- Val. M
Importar
Limpar Dados

Processamento

Realizar classificação

Probabilidade a priori de óbito
0,5

Altera probabilidade a priori

☒ Covariação Vivos
☒ Covariação Mortos
☒ Inversão Vivos
☒ Inversão Mortos

☒ Cholesky Vivos
☒ Cholesky Mortos
☒ Determinantes
☒ Resultados

	Vivos	Óbitos
Dos Vivos	16886	1658
Dos Mortos	83	143

Salvar classificação

Taxa aparente de erro 9,27%

16:13:48 Importando dados
--> Progresso: 226
16:13:58 Calculando vetor médio
16:13:58 --> População vivos : 18544
16:13:58 --> População mortos : 226

FONTE: O AUTOR

Resultados com probabilidades a priori diferentes: (óbito igual a 1,2% e sobrevivência igual a 98,8%):

FIGURA 4.9 – RESULTADOS COM AMOSTRA 01 – PROBABILIDADE A PRIORI DIFERENTES

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

	Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
▶	4	25	5	2	5	1	2	4	0	0	2	3100	0			
	1	34	2	4	5	1	2	4	0	8	10	3330	0			
	1	18	2	4	5	1	1	4	0	9	10	3160	0			

Processamento

Realizar classificação

Probabilidade a priori de óbito: 0,012

Alterar probabilidade a priori

Salvar classificação

Taxa aparente de erro: 4,60%

Importar: 4- Val. M

Importar

Limpar Dados

☒ Covariância Vivos
☒ Covariância Mortos
☒ Inversão Vivos
☒ Inversão Mortos

☒ Cholesky Vivos
☒ Cholesky Mortos
☒ Determinantes
☒ Resultados

	Vivos	Óbitos
Dos Vivos	7112	306
Dos Mortos	40	50

14:15:43 Calculando subtração das inversas das covariâncias
 14:15:43 Obtendo resultados para vivos
 14:15:43 --> Ln(priori/m/prioriv): 0
 14:15:44 Obtendo resultados para mortos
 14:15:44 Fim da classificação

FONTE: O AUTOR

FIGURA 4.10 – RESULTADOS COM AMOSTRA 02 – PROBABILIDADE A PRIORI DIFERENTES

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

	Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
	4	25	1	3	5	1	1	4	1	8	9	3110	9			
	4	16	1	4	2	1	1	2	0	1	8	600	1			
▶	4	26	1	4	4	1	1	3	0	6	8	1700	5			

Processamento

Realizar classificação

Probabilidade a priori de óbito: 0,012

Alterar probabilidade a priori

Salvar classificação

Taxa aparente de erro: 4,72%

Importar: 4- Val. M

Importar

Limpar Dados

☒ Covariância Vivos
☒ Covariância Mortos
☒ Inversão Vivos
☒ Inversão Mortos

☒ Cholesky Vivos
☒ Cholesky Mortos
☒ Determinantes
☒ Resultados

	Vivos	Óbitos
Dos Vivos	7110	308
Dos Mortos	47	43

--> Progresso: 90
 14:44:13 Calculando vetor médio
 14:44:13 --> População vivos : 11126
 14:44:14 --> População mortos : 136
 14:44:14 Calculando covariância para vivos

FONTE: O AUTOR

FIGURA 4.11 – RESULTADOS COM AMOSTRA 03 – PROBABILIDADE A PRIORI DIFERENTES

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	21	5	2	5	1	2	3	1	5	8	2470	4			
4	20	1	4	2	1	1	1	0	4	7	800	5			
4	20	1	4	5	1	2	3	0	9	10	3350	5			

Processamento

Realizar classificação

Probabilidade a priori de óbito: 0,012

Alterar probabilidade a priori

Salvar classificação

☒ Covariância Vivos
☒ Covariância Mortos
☒ Inversão Vivos
☒ Inversão Mortos

☒ Cholesky Vivos
☒ Cholesky Mortos
☒ Determinantes
☒ Resultados

	Vivos	Óbitos
Dos Vivos	7110	298
Dos Mortos	42	48

Taxa aparente de erro: 4,53%

15:00:59 Importando dados
 --> Progresso: 90
 15:01:06 Calculando vetor médio
 15:01:06 --> População vivos : 11136
 15:01:06 --> População mortos : 136

FONTE: O AUTOR

FIGURA 4.12 – RESULTADOS COM AMOSTRA 04 – PROBABILIDADE A PRIORI DIFERENTES

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	28	1	3	5	1	1	4	1	7	8	4040	0			
4	37	1	3	4	1	1	3	0	7	8	2920	4			
4	25	5	2	5	1	2	4	0	0	2	3100	0			

Processamento

Realizar classificação

Probabilidade a priori de óbito: 0,012

Alterar probabilidade a priori

Salvar classificação

☒ Covariância Vivos
☒ Covariância Mortos
☒ Inversão Vivos
☒ Inversão Mortos

☒ Cholesky Vivos
☒ Cholesky Mortos
☒ Determinantes
☒ Resultados

	Vivos	Óbitos
Dos Vivos	7135	283
Dos Mortos	43	47

Taxa aparente de erro: 4,34%

15:13:39 Importando dados
 --> Progresso: 90
 15:13:45 Calculando vetor médio
 15:13:46 --> População vivos : 11126
 15:13:46 --> População mortos : 136

FONTE: O AUTOR

FIGURA 4.13 – RESULTADOS COM AMOSTRA 05 – PROBABILIDADE A PRIORI DIFERENTES

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	34	5	3	2	1	1	2	0	4	1	850	2			
4	19	1	3	4	1	1	4	0	3	5	1950	0			
4	18	1	4	2	1	1	1	1	4	6	500	6			

Importar
4- Val. M
Importar
Limpar Dados

Processamento

Realizar classificação

Probabilidade a priori de óbito
0,012

Altera probabilidade a priori

Salvar classificação

☒ Covariância Vivos
☒ Covariância Mortos
☒ Inversão Vivos
☒ Inversão Mortos

☒ Cholesky Vivos
☒ Cholesky Mortos
☒ Determinantes
☒ Resultados

	Vivos	Óbitos
Dos Vivos	7127	291
Dos Mortos	41	49

Taxa aparente de erro 4,42%

15:24:37 Importando dados
--> Progresso: 90
15:24:45 Calculando vetor médio
15:24:45 --> População vivos : 11126
15:24:45 --> População mortos : 136

FONTE: O AUTOR

Considerando-se todos os dados disponíveis tanto para a construção da função de reconhecimento e classificação quanto para a validação da mesma, obtiveram-se os seguintes resultados:

FIGURA 4.14 – RESULTADOS COM UTILIZAÇÃO DE TODAS AS OBSERVAÇÕES – PROBABILIDADES A PRIORI DIFERENTES

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	25	1	3	5	1	1	4	1	8	9	3110	9			
4	16	1	4	2	1	1	2	0	1	8	600	1			
4	26	1	4	4	1	1	3	0	6	8	1700	5			

Importar
4- Val. M
Importar
Limpar Dados

Processamento

Realizar classificação

Probabilidade a priori de óbito
0,012

Altera probabilidade a priori

Salvar classificação

☒ Covariância Vivos
☒ Covariância Mortos
☒ Inversão Vivos
☒ Inversão Mortos

☒ Cholesky Vivos
☒ Cholesky Mortos
☒ Determinantes
☒ Resultados

	Vivos	Óbitos
Dos Vivos	17785	759
Dos Mortos	108	118

Taxa aparente de erro 4,61%

16:13:48 Importando dados
--> Progresso: 226
16:13:58 Calculando vetor médio
16:13:58 --> População vivos : 18544
16:13:58 --> População mortos : 226

FONTE: O AUTOR

Considerando-se todos os dados disponíveis tanto para a construção da função de reconhecimento e classificação quanto para a validação da mesma, com exclusão de variáveis e probabilidades a priori iguais, obtiveram-se os seguintes resultados:

FIGURA 4.15 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL IDADE DA MÃE

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	1	3	5	1	1	4	1	8	9	3110	9				
4	1	4	2	1	1	2	0	1	8	600	1				
4	1	4	4	1	1	3	0	6	8	1700	5				

Processamento

Realizar classificação

Probabilidade a priori de óbito: 0,5

Altera probabilidade a priori **Salvar classificação**

☒ Covariância Vivos ☒ Cholesky Vivos
☒ Covariância Mortos ☒ Cholesky Mortos
☒ Inversão Vivos ☒ Determinantes
☒ Inversão Mortos ☒ Resultados

	Vivos	Óbitos
Dos Vivos	16880	1664
Dos Mortos	82	144

Taxa aparente de erro: 9,30%

20:25:39 Importando dados
 --> Progresso: 226
 20:25:44 Calculando vetor médio
 20:25:44 --> População vivos : 16544
 20:25:44 --> População mortos : 226

FONTE: O AUTOR

FIGURA 4.16 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL ESTADO CIVIL DA MÃE

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	25	3	5	1	1	4	1	8	9	3110	9				
4	16	4	2	1	1	2	0	1	8	600	1				
4	26	4	4	1	1	3	0	6	8	1700	5				

Processamento

Realizar classificação

Probabilidade a priori de óbito: 0,5

Altera probabilidade a priori **Salvar classificação**

☒ Covariância Vivos ☒ Cholesky Vivos
☒ Covariância Mortos ☒ Cholesky Mortos
☒ Inversão Vivos ☒ Determinantes
☒ Inversão Mortos ☒ Resultados

	Vivos	Óbitos
Dos Vivos	16909	1635
Dos Mortos	86	140

Taxa aparente de erro: 9,16%

20:30:56 Importando dados
 --> Progresso: 226
 20:31:00 Calculando vetor médio
 20:31:01 --> População vivos : 16544
 20:31:01 --> População mortos : 226

FONTE: O AUTOR

FIGURA 4.17 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL ESCOLARIDADE DA MÃE

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	25	1	5	1	1	4	1	8	9	3110	9				
4	16	1	2	1	1	2	0	1	8	600	1				
4	26	1	4	1	1	3	0	6	8	1700	5				

Processamento

Realizar classificação

Probabilidade a priori de óbito: 0,5

Altera probabilidade a priori

Salvar classificação

Taxa aparente de erro: 9,49%

Importar: 4- Val. M, Importar, Limpar Dados

20:36:22 Importando dados
 --> Progresso: 226
 20:36:30 Calculando vetor médio
 20:36:30 --> População vivos : 18544
 20:36:30 --> População mortos : 226

FONTE: O AUTOR

FIGURA 4.18 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL DURAÇÃO DA GESTAÇÃO

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	25	1	3	1	1	4	1	8	9	3110	9				
4	16	1	4	1	1	2	0	1	8	600	1				
4	26	1	4	1	1	3	0	6	8	1700	5				

Processamento

Realizar classificação

Probabilidade a priori de óbito: 0,5

Altera probabilidade a priori

Salvar classificação

Taxa aparente de erro: 8,01%

Importar: 4- Val. M, Importar, Limpar Dados

20:47:32 Importando dados
 --> Progresso: 226
 20:47:37 Calculando vetor médio
 20:47:37 --> População vivos : 18544
 20:47:37 --> População mortos : 226

FONTE: O AUTOR

FIGURA 4.19 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL TIPO DE GRAVIDEZ

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	25	1	3	5	1	4	1	8	9	3110	9				
4	16	1	4	2	1	2	0	1	8	600	1				
4	26	1	4	4	1	3	0	6	8	1700	5				

Processamento

Probabilidade a priori de óbito: 0,5

☒ Covariância Vivos
☒ Covariância Mortos
☒ Inversão Vivos
☒ Inversão Mortos

☒ Cholesky Vivos
☒ Cholesky Mortos
☒ Determinantes
☒ Resultados

	Vivos	Óbitos
Dos Vivos	16891	1653
Dos Mortos	80	146

Taxa aparente de erro: 9,23%

20:53:13 Importando dados
 --> Progresso: 226
 20:53:19 Calculando vetor médio
 20:53:19 --> População vivos: 18544
 20:53:20 --> População mortos: 226

FONTE: O AUTOR

FIGURA 4.20 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL TIPO DE PARTO

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	25	1	3	5	1	4	1	8	9	3110	9				
4	16	1	4	2	1	2	0	1	8	600	1				
4	26	1	4	4	1	3	0	6	8	1700	5				

Processamento

Probabilidade a priori de óbito: 0,5

☒ Covariância Vivos
☒ Covariância Mortos
☒ Inversão Vivos
☒ Inversão Mortos

☒ Cholesky Vivos
☒ Cholesky Mortos
☒ Determinantes
☒ Resultados

	Vivos	Óbitos
Dos Vivos	16868	1676
Dos Mortos	84	142

Taxa aparente de erro: 9,37%

20:57:49 Importando dados
 --> Progresso: 226
 20:58:01 Calculando vetor médio
 20:58:01 --> População vivos: 18544
 20:58:02 --> População mortos: 226

FONTE: O AUTOR

FIGURA 4.21 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL NÚMERO DE CONSULTAS PRÉ-NATAL

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	25	1	3	5	1	1	1	8	9	3110	9				
4	16	1	4	2	1	1	0	1	8	600	1				
4	26	1	4	4	1	1	0	6	8	1700	5				

Importar
4- Val. M
Importar
Limpar Dados

Processamento

Realizar classificação

Probabilidade a priori de óbito
0,5

Alterar probabilidade a priori

Salvar classificação

Taxa aparente de erro 9,13%

21:02:40 Importando dados
--> Progresso: 226
21:02:46 Calculando vetor médio
21:02:47 --> População vivos : 18544
21:02:47 --> População mortos : 226

FONTE: O AUTOR

FIGURA 4.22 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL SEXO

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	25	1	3	5	1	1	4	8	9	3110	9				
4	16	1	4	2	1	1	2	1	8	600	1				
4	26	1	4	4	1	1	3	6	8	1700	5				

Importar
4- Val. M
Importar
Limpar Dados

Processamento

Realizar classificação

Probabilidade a priori de óbito
0,5

Alterar probabilidade a priori

Salvar classificação

Taxa aparente de erro 9,28%

21:09:11 Calculando covariância para vivos
21:09:31 Calculando covariância para mortos
21:09:31 Calculando inversão para vivos
21:09:31 Calculando inversão para mortos
21:09:31 Calculando decomposição de cholesky para vivos

FONTE: O AUTOR

FIGURA 4.23 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL APGAR 1º MINUTO

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	25	1	3	5	1	1	4	1	9	3110	9				
4	16	1	4	2	1	1	2	0	8	600	1				
4	26	1	4	4	1	1	3	0	8	1700	5				

Processamento

Realizar classificação

Probabilidade a priori de óbito: 0,5

☐ Covariância Vivos ☒ Cholesky Vivos

☒ Covariância Mortos ☒ Cholesky Mortos

☒ Inversão Vivos ☒ Determinantes

☒ Inversão Mortos ☒ Resultados

	Vivos	Óbitos
Dos Vivos	16944	1600
Dos Mortos	84	142

Taxa aparente de erro: 8,97%

21:15:10 Importando dados
 --> Progresso: 226
 21:15:19 Calculando vetor médio
 21:15:20 --> População vivos : 18544
 21:15:20 --> População mortos : 226

FONTE: O AUTOR

FIGURA 4.24 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL APGAR 5º MINUTO

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	25	1	3	5	1	1	4	1	8	3110	9				
4	16	1	4	2	1	1	2	0	1	600	1				
4	26	1	4	4	1	1	3	0	6	1700	5				

Processamento

Realizar classificação

Probabilidade a priori de óbito: 0,5

☒ Covariância Vivos ☒ Cholesky Vivos

☒ Covariância Mortos ☒ Cholesky Mortos

☒ Inversão Vivos ☒ Determinantes

☒ Inversão Mortos ☒ Resultados

	Vivos	Óbitos
Dos Vivos	16777	1767
Dos Mortos	81	145

Taxa aparente de erro: 9,84%

21:20:10 Importando dados
 --> Progresso: 226
 21:20:20 Calculando vetor médio
 21:20:20 --> População vivos : 18544
 21:20:21 --> População mortos : 226

FONTE: O AUTOR

FIGURA 4.25 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL PESO

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	25	1	3	5	1	1	4	1	8	9	9				
4	16	1	4	2	1	1	2	0	1	8	1				
4	26	1	4	4	1	1	3	0	6	8	5				

Importar
4- Val. M
Importar
Limpar Dados

Processamento

Realizar classificação

Probabilidade a priori de óbito
0,5

Alterar probabilidade a priori

Salvar classificação

☒ Covariância Vivos
☒ Covariância Mortos
☒ Inversão Vivos
☒ Inversão Mortos

☒ Cholesky Vivos
☒ Cholesky Mortos
☒ Determinantes
☒ Resultados

	Vivos	Óbitos
Dos Vivos	16834	1710
Dos Mortos	82	144

Taxa aparente de erro 9,54%

21:24:50 Importando dados
--> Progresso: 226
21:24:56 Calculando vetor médio
21:24:56 --> População vivos : 18544
21:24:57 --> População mortos : 226

FONTE: O AUTOR

FIGURA 4.26 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL NÚMERO DE GESTAÇÕES ANTERIORES

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	25	1	3	5	1	1	4	1	8	9	3110				
4	16	1	4	2	1	1	2	0	1	8	600				
4	26	1	4	4	1	1	3	0	6	8	1700				

Importar
4- Val. M
Importar
Limpar Dados

Processamento

Realizar classificação

Probabilidade a priori de óbito
0,5

Alterar probabilidade a priori

Salvar classificação

☒ Covariância Vivos
☒ Covariância Mortos
☒ Inversão Vivos
☒ Inversão Mortos

☒ Cholesky Vivos
☒ Cholesky Mortos
☒ Determinantes
☒ Resultados

	Vivos	Óbitos
Dos Vivos	16912	1632
Dos Mortos	87	139

Taxa aparente de erro 9,15%

21:28:46 Importando dados
--> Progresso: 226
21:28:57 Calculando vetor médio
21:28:57 --> População vivos : 18544
21:28:58 --> População mortos : 226

FONTE: O AUTOR

FIGURA 4.27 – RESULTADOS OBTIDOS COM A UTILIZAÇÃO DAS VARIÁVEIS APGAR 1, APGAR 5 E PESO

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	8	9	3110												
4	1	8	600												
4	6	8	1700												

Importar: 4- Val. M

Importar

Limpar Dados

Processamento

Realizar classificação

Probabilidade a priori de óbito: 0,5

Alterar probabilidade a priori

Salvar classificação

☒ Covariância Vivos
☒ Covariância Mortos
☒ Inversão Vivos
☒ Inversão Mortos

☒ Cholesky Vivos
☒ Cholesky Mortos
☒ Determinantes
☒ Resultados

	Vivos	Óbitos
Dos Vivos	17293	1251
Dos Mortos	100	126

Taxa aparente de erro: 7,19%

21:37:01 Importando dados
 --> Progresso: 226
 21:37:05 Calculando vetor médio
 21:37:06 --> População vivos : 18544
 21:37:06 --> População mortos : 226

FONTE: O AUTOR

FIGURA 4.28 – RESULTADOS OBTIDOS COM A EXCLUSÃO DAS VARIÁVEIS APGAR 1, APGAR 5 E PESO

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	25	1	3	5	1	1	4	1	9						
4	16	1	4	2	1	1	2	0	1						
4	26	1	4	4	1	1	3	0	5						

Importar: 4- Val. M

Importar

Limpar Dados

Processamento

Realizar classificação

Probabilidade a priori de óbito: 0,5

Alterar probabilidade a priori

Salvar classificação

☒ Covariância Vivos
☒ Covariância Mortos
☒ Inversão Vivos
☒ Inversão Mortos

☒ Cholesky Vivos
☒ Cholesky Mortos
☒ Determinantes
☒ Resultados

	Vivos	Óbitos
Dos Vivos	16848	1696
Dos Mortos	103	123

Taxa aparente de erro: 9,58%

21:43:13 Importando dados
 --> Progresso: 226
 21:43:18 Calculando vetor médio
 21:43:18 --> População vivos : 18544
 21:43:19 --> População mortos : 226

FONTE: O AUTOR

FIGURA 4.29 – RESULTADOS OBTIDOS COM A UTILIZAÇÃO DAS VARIÁVEIS IDADE DA MÃE, APGAR 1, APGAR 5 E PESO

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	25	8	9	3110											
4	16	1	8	600											
4	26	6	8	1700											

Processamento

Realizar classificação

Probabilidade a priori de óbito: 0,5

Altera probabilidade a priori

Salvar classificação

Covariação Vivos ☒ Cholesky Vivos ☒
 Covariação Mortos ☒ Cholesky Mortos ☒
 Inversão Vivos ☒ Determinantes ☒
 Inversão Mortos ☒ Resultados ☒

	Vivos	Óbitos
Dos Vivos	17275	1269
Dos Mortos	100	126

Taxa aparente de erro: 7,29%

21:57:46 Importando dados
 --> Progresso: 226
 21:57:53 Calculando vetor médio
 21:57:53 --> População vivos: 18544
 21:57:54 --> População mortos: 226

FONTE: O AUTOR

FIGURA 4.30 – RESULTADOS OBTIDOS COM A EXCLUSÃO DAS VARIÁVEIS IDADE DA MÃE, APGAR 1, APGAR 5 E PESO

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	1	4	4	1	1	3	0	5							
1	1	3	5	1	2	4	1	1							
1	5	3	5	1	1	4	0	1							

Processamento

Realizar classificação

Probabilidade a priori de óbito: 0,5

Altera probabilidade a priori

Salvar classificação

Covariação Vivos ☒ Cholesky Vivos ☒
 Covariação Mortos ☒ Cholesky Mortos ☒
 Inversão Vivos ☒ Determinantes ☒
 Inversão Mortos ☒ Resultados ☒

	Vivos	Óbitos
Dos Vivos	16873	1671
Dos Mortos	105	121

Taxa aparente de erro: 9,46%

10:20:33 Importando dados
 --> Progresso: 226
 10:20:41 Calculando vetor médio
 10:20:41 --> População vivos: 18544
 10:20:41 --> População mortos: 226

FONTE: O AUTOR

Considerando-se todos os dados disponíveis tanto para a construção da função de reconhecimento e classificação quanto para a validação da mesma, com exclusão de variáveis e probabilidade a priori diferentes (1,2% e 98,8%), obtiveram-se os seguintes resultados:

FIGURA 4.31 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL IDADE DA MÃE

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	1	3	5	1	1	4	1	8	9	3110	9				
4	1	4	2	1	1	2	0	1	8	600	1				
4	1	4	4	1	1	3	0	6	8	1700	5				

Processamento

Realizar classificação

Probabilidade a priori de óbito: 0,012

Alterar probabilidade a priori

Salvar classificação

Taxa aparente de erro: 4,60%

Importar: 4- Val. M

Importar

Limpar Dados

12:07:45 Importando dados
--> Progresso: 226
12:07:55 Calculando vetor médio
12:07:55 --> População vivos : 18544
12:07:56 --> População mortos : 226

FONTE: O AUTOR

FIGURA 4.32 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL ESTADO CIVIL DA MÃE

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	25	3	5	1	1	4	1	8	9	3110	9				
4	16	4	2	1	1	2	0	1	8	600	1				
4	26	4	4	1	1	3	0	6	8	1700	5				

Processamento

Realizar classificação

Probabilidade a priori de óbito: 0,012

Alterar probabilidade a priori

Salvar classificação

Taxa aparente de erro: 4,59%

Importar: 4- Val. M

Importar

Limpar Dados

11:59:49 Importando dados
--> Progresso: 226
11:59:56 Calculando vetor médio
11:59:56 --> População vivos : 18544
11:59:57 --> População mortos : 226

FONTE: O AUTOR

FIGURA 4.33 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL ESCOLARIDADE DA MÃE

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	25	1	5	1	1	4	1	8	9	3110	9				
4	16	1	2	1	1	2	0	1	8	600	1				
4	26	1	4	1	1	3	0	6	8	1700	5				

Importar
4- Val. M
Importar
Limpar Dados

Processamento

Realizar classificação

Probabilidade a priori de óbito
0,012

Alterar probabilidade a priori

Salvar classificação

Taxa aparente de erro 4,62%

11:55:04 Importando dados
--> Progresso: 226
11:55:15 Calculando vetor médio
11:55:15 --> População vivos : 18544
11:55:16 --> População mortos : 226

FONTE: O AUTOR

FIGURA 4.34 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL DURAÇÃO DA GESTAÇÃO

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	25	1	3	1	1	4	1	8	9	3110	9				
4	16	1	4	1	1	2	0	1	8	600	1				
4	26	1	4	1	1	3	0	6	8	1700	5				

Importar
4- Val. M
Importar
Limpar Dados

Processamento

Realizar classificação

Probabilidade a priori de óbito
0,012

Alterar probabilidade a priori

Salvar classificação

Taxa aparente de erro 4,28%

11:49:26 Importando dados
--> Progresso: 226
11:49:34 Calculando vetor médio
11:49:34 --> População vivos : 18544
11:49:34 --> População mortos : 226

FONTE: O AUTOR

FIGURA 4.35 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL TIPO DE GRAVIDEZ

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	25	1	3	5	1	4	1	8	9	3110	9				
4	16	1	4	2	1	2	0	1	8	600	1				
4	26	1	4	4	1	3	0	6	8	1700	5				

Importar
4- Val. M
Importar
Limpar Dados

Processamento

Realizar classificação

Probabilidade a priori de óbito
0,012

Altera probabilidade a priori

Salvar classificação

Taxa aparente de erro 3,30%

11:25:48 Importando dados
--> Progresso: 226
11:25:56 Calculando vetor médio
11:25:56 --> População vivos : 18544
11:25:56 --> População mortos : 226

FONTE: O AUTOR

FIGURA 4.36 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL TIPO DE PARTO

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	25	1	3	5	1	4	1	8	9	3110	9				
4	16	1	4	2	1	2	0	1	8	600	1				
4	26	1	4	4	1	3	0	6	8	1700	5				

Importar
4- Val. M
Importar
Limpar Dados

Processamento

Realizar classificação

Probabilidade a priori de óbito
0,012

Altera probabilidade a priori

Salvar classificação

Taxa aparente de erro 4,64%

11:21:10 Importando dados
--> Progresso: 226
11:21:17 Calculando vetor médio
11:21:18 --> População vivos : 18544
11:21:18 --> População mortos : 226

FONTE: O AUTOR

FIGURA 4.37 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL NÚMERO DE CONSULTAS PRÉ-NATAL

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	25	1	3	5	1	1	1	8	9	3110	9				
4	16	1	4	2	1	1	0	1	8	600	1				
4	26	1	4	4	1	1	0	6	8	1700	5				

Importar
4- Val. M
Importar
Limpar Dados

Processamento

Realizar classificação

Probabilidade a priori de óbito
0,012

Altera probabilidade a priori

Salvar classificação

Taxa aparente de erro 4,54%

11:16:33 Importando dados
--> Progresso: 226
11:16:43 Calculando vetor médio
11:16:44 --> População vivos : 18544
11:16:44 --> População mortos : 226

FONTE: O AUTOR

FIGURA 4.38 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL SEXO

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	25	1	3	5	1	1	4	8	9	3110	9				
4	16	1	4	2	1	1	2	1	8	600	1				
4	26	1	4	4	1	1	3	6	8	1700	5				

Importar
4- Val. M
Importar
Limpar Dados

Processamento

Realizar classificação

Probabilidade a priori de óbito
0,012

Altera probabilidade a priori

Salvar classificação

Taxa aparente de erro 4,64%

11:11:41 Importando dados
--> Progresso: 226
11:11:54 Calculando vetor médio
11:11:54 --> População vivos : 18544
11:11:54 --> População mortos : 226

FONTE: O AUTOR

FIGURA 4.39 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL APGAR 1º MINUTO

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	25	1	3	5	1	1	4	1	9	3110	9				
4	16	1	4	2	1	1	2	0	8	600	1				
4	26	1	4	4	1	1	3	0	8	1700	5				

Processamento

Realizar classificação

Probabilidade a priori de óbito
0,012

Alterar probabilidade a priori

Salvar classificação

☒ Covariância Vivos
☒ Covariância Mortos
☒ Inversão Vivos
☒ Inversão Mortos

☒ Cholesky Vivos
☒ Cholesky Mortos
☒ Determinantes
☒ Resultados

	Vivos	Óbitos
Dos Vivos	17888	656
Dos Mortos	109	117

Taxa aparente de erro 4,07%

11:05:42 Importando dados
--> Progresso: 226
11:05:50 Calculando vetor médio
11:05:50 --> População vivos : 18544
11:05:51 --> População mortos : 226

FONTE: O AUTOR

FIGURA 4.40 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL APGAR 5.º MINUTO

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	25	1	3	5	1	1	4	1	8	3110	9				
4	16	1	4	2	1	1	2	0	1	600	1				
4	26	1	4	4	1	1	3	0	6	1700	5				

Processamento

Realizar classificação

Probabilidade a priori de óbito
0,012

Alterar probabilidade a priori

Salvar classificação

☒ Covariância Vivos
☒ Covariância Mortos
☒ Inversão Vivos
☒ Inversão Mortos

☒ Cholesky Vivos
☒ Cholesky Mortos
☒ Determinantes
☒ Resultados

	Vivos	Óbitos
Dos Vivos	17819	725
Dos Mortos	110	116

Taxa aparente de erro 4,44%

11:00:08 Importando dados
--> Progresso: 226
11:00:19 Calculando vetor médio
11:00:19 --> População vivos : 18544
11:00:20 --> População mortos : 226

FONTE: O AUTOR

FIGURA 4.41 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL PESO

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	25	1	3	5	1	1	4	1	8	9	9				
4	16	1	4	2	1	1	2	0	1	8	1				
4	26	1	4	4	1	1	3	0	6	8	5				

Importar
4- Val. M
Importar
Limpar Dados

Processamento

Realizar classificação

Probabilidade a priori de óbito
0,012

Alterar probabilidade a priori

Salvar classificação

Taxa aparente de erro 4,60%

10:54:36 Importando dados
--> Progresso: 226
10:54:47 Calculando vetor médio
10:54:47 --> População vivos : 18544
10:54:48 --> População mortos : 226

FONTE: O AUTOR

FIGURA 4.42 – RESULTADOS OBTIDOS COM A EXCLUSÃO DA VARIÁVEL NÚMERO DE GESTAÇÕES ANTERIORES

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	25	1	3	5	1	1	4	1	8	9	3110				
4	16	1	4	2	1	1	2	0	1	8	600				
4	26	1	4	4	1	1	3	0	6	8	1700				

Importar
4- Val. M
Importar
Limpar Dados

Processamento

Realizar classificação

Probabilidade a priori de óbito
0,012

Alterar probabilidade a priori

Salvar classificação

Taxa aparente de erro 4,54%

10:48:21 Importando dados
--> Progresso: 226
10:48:30 Calculando vetor médio
10:48:30 --> População vivos : 18544
10:48:31 --> População mortos : 226

FONTE: O AUTOR

FIGURA 4.43 – RESULTADOS OBTIDOS COM A UTILIZAÇÃO DAS VARIÁVEIS APGAR 1, APGAR 5 E PESO

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	8	9	3110												
4	1	8	600												
4	6	8	1700												

Processamento

Realizar classificação

Probabilidade a priori de óbito: 0,012

Alterar probabilidade a priori

Salvar classificação

☒ Covariação Vivos
☒ Covariação Mortos
☒ Inversão Vivos
☒ Inversão Mortos

☒ Cholesky Vivos
☒ Cholesky Mortos
☒ Determinantes
☒ Resultados

	Vivos	Óbitos
Dos Vivos	18146	398
Dos Mortos	129	97

Taxa aparente de erro: 2,80%

10:43:49 Importando dados
 --> Progresso: 226
 10:43:57 Calculando vetor médio
 10:43:57 --> População vivos : 18544
 10:43:57 --> População mortos : 226

FONTE: O AUTOR

FIGURA 4.44 – RESULTADOS OBTIDOS COM A EXCLUSÃO DAS VARIÁVEIS APGAR 1, APGAR 5 E PESO

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	25	1	3	5	1	1	4	1	9						
4	16	1	4	2	1	1	2	0	1						
4	26	1	4	4	1	1	3	0	5						

Processamento

Realizar classificação

Probabilidade a priori de óbito: 0,012

Alterar probabilidade a priori

Salvar classificação

☒ Covariação Vivos
☒ Covariação Mortos
☒ Inversão Vivos
☒ Inversão Mortos

☒ Cholesky Vivos
☒ Cholesky Mortos
☒ Determinantes
☒ Resultados

	Vivos	Óbitos
Dos Vivos	17984	560
Dos Mortos	146	80

Taxa aparente de erro: 3,76%

10:39:03 Importando dados
 --> Progresso: 226
 10:39:13 Calculando vetor médio
 10:39:13 --> População vivos : 18544
 10:39:13 --> População mortos : 226

FONTE: O AUTOR

FIGURA 4.45 – RESULTADOS OBTIDOS COM A UTILIZAÇÃO DAS VARIÁVEIS IDADE DA MÃE, APGAR 1, APGAR 5 E PESO

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	26	6	8	1700											
1	14	9	10	3050											
1	14	10	10	3100											

Processamento

Realizar classificação

Probabilidade a priori de óbito: 0,012

Alterar probabilidade a priori

Salvar classificação

Covariância Vivos ☒ Cholesky Vivos ☒
 Covariância Mortos ☒ Cholesky Mortos ☒
 Inversão Vivos ☒ Determinantes ☒
 Inversão Mortos ☒ Resultados ☒

	Vivos	Óbitos
Dos Vivos	18139	405
Dos Mortos	128	98

Taxa aparente de erro: 2,83%

10:33:23 Importando dados
 --> Progresso: 226
 10:33:33 Calculando vetor médio
 10:33:33 --> População vivos : 18544
 10:33:33 --> População mortos : 226

FONTE: O AUTOR

FIGURA 4.46 – RESULTADOS OBTIDOS COM A EXCLUSÃO DAS VARIÁVEIS IDADE DA MÃE, APGAR 1, APGAR 5 E PESO

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Gr.	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12	C13	C14	C15
4	1	4	4	1	1	3	0	5							
1	1	3	5	1	2	4	1	1							
1	5	3	5	1	1	4	0	1							

Processamento

Realizar classificação

Probabilidade a priori de óbito: 0,012

Alterar probabilidade a priori

Salvar classificação

Covariância Vivos ☒ Cholesky Vivos ☒
 Covariância Mortos ☒ Cholesky Mortos ☒
 Inversão Vivos ☒ Determinantes ☒
 Inversão Mortos ☒ Resultados ☒

	Vivos	Óbitos
Dos Vivos	17984	560
Dos Mortos	146	80

Taxa aparente de erro: 3,76%

10:29:02 Lendo classificação
 10:29:26 Calculando vetor médio
 10:29:26 --> População vivos : 18544
 10:29:26 --> População mortos : 226
 10:29:26 Calculando covariância para vivos

FONTE: O AUTOR

4.4 AVALIAÇÃO DO MODELO

4.4.1 Análise dos resultados sem exclusão de variáveis

Comparando-se as tabelas 4.1 e 4.2 a seguir, observou-se, em todos os casos, que o desempenho da função de reconhecimento e classificação foi melhor quando se considerou as probabilidades a priori observadas nos grupos estudados, ou seja, probabilidade de óbito igual a 1,2% e probabilidade de sobrevivência igual a 98,8%, pois as taxas de erro foram menores. Este resultado foi influenciado pela grande diferença na quantidade de dados existente entre os dois grupos de observações. Existe uma tendência natural em classificar melhor o grupo com maior quantidade de observações [HAIR *et. al.*, 2005], uma vez que existe maior informação sobre este grupo.

TABELA 4.1 – ANÁLISE DOS RESULTADOS SEM EXCLUSÃO DE VARIÁVEIS – PROBABILIDADE A PRIORI DIFERENTES

AMOSTRA N°	TAXA DE ERRO	PORCENTAGEM DE INDIVÍDUOS CORRETAMENTE CLASSIFICADOS	
		SOBREVIVÊNCIA	ÓBITO
01	4,60 %	95,87 %	55,56 %
02	4,72 %	95,85 %	47,78 %
03	4,53 %	95,85 %	53,33 %
04	4,34 %	96,18 %	52,22 %
05	4,42 %	96,08 %	54,44 %
Usando 100% dos dados	4,61 %	95,91 %	52,21 %

FONTE: O AUTOR

Porém, deve-se considerar neste estudo, que a classificação correta de crianças com probabilidade de óbito deve ser essencial e, portanto, os melhores resultados, segundo esse raciocínio, foram obtidos considerando as probabilidades a priori iguais a 50%.

Observou-se, também, que nas cinco situações em que foram tomadas amostras para a construção da função de reconhecimento e classificação com 60% dos dados e amostras de validação com os 40% restantes dos dados (conforme sugere HAIR *et. al.* [2005]), os resultados de classificações corretas para as amostras de validação apresentaram

certa estabilidade para o grupo de sobrevivência (em torno de 91%), enquanto que para o grupo de óbito essa estabilidade não se observou (variou de 57,78% a em torno de 65,56%).

TABELA 4.2 – ANÁLISE DOS RESULTADOS SEM EXCLUSÃO DE VARIÁVEIS - PROBABILIDADES A PRIORI IGUAIS

AMOSTRA N°	TAXA DE ERRO	PORCENTAGEM DE INDIVÍDUOS CORRETAMENTE CLASSIFICADOS	
		SOBREVIVÊNCIA	ÓBITO
01	8,73 %	91,61 %	62,22 %
02	9,33 %	91,06 %	57,78 %
03	8,92 %	91,26 %	65,56 %
04	8,77 %	91,57 %	62,22 %
05	8,55 %	91,79 %	63,33 %
Usando 100% dos dados	9,27 %	91,06 %	63,27 %

FONTE: O AUTOR

A utilização da função discriminante quadrática apresentou resultados satisfatórios atingindo uma classificação correta de 63,27% dos indivíduos que chegaram a óbito e 91,06% para a população de sobrevivência quando 100% dos dados foram usados para a construção da função de reconhecimento e classificação e para a validação dos resultados (tabela 4.2).

4.4.2 Análise dos resultados com exclusão de variáveis

Com probabilidades a priori, de sobrevivência e de óbito, iguais a 50% e excluindo as variáveis estudadas, uma a uma, e analisando o desempenho da função de reconhecimento e classificação, os resultados foram:

TABELA 4.3 – ANÁLISE DOS RESULTADOS EXCLUINDO AS VARIÁVEIS UMA A UMA

VARIÁVEL EXCLUÍDA	APER	PORCENTAGEM DE INDIVÍDUOS CORRETAMENTE CLASSIFICADOS	
		SOBREVIVÊNCIA	ÓBITO
Nenhuma	9,27 %	91,06 %	63,27 %
Idade da Mãe	9,30 %	91,03 %	63,72 %
Estado Civil da Mãe	9,16 %	91,18 %	61,95 %
Duração da Gestação	8,01 %	92,35 %	61,95 %
Tipo de Parto	9,37 %	90,96 %	62,83 %

VARIÁVEL EXCLUÍDA	APER	PORCENTAGEM DE INDIVÍDUOS CORRETAMENTE CLASSIFICADOS	
		SOBREVIVÊNCIA	ÓBITO
Tipo de Gravidez	9,23 %	91,09 %	64,60 %
Número de Consultas de Pré-Natal	9,13 %	91,20 %	62,83 %
Sexo	9,28 %	91,05 %	62,83 %
Apgar no 1º Minuto	8,97 %	91,37 %	62,83 %
Apgar no 5º Minuto	9,84 %	90,47 %	64,16 %
Peso	9,54 %	90,79 %	63,72 %
Número de Gestações Anteriores	9,15 %	91,20 %	61,50 %

FONTE: O AUTOR

Observou-se que a retirada individual de cada uma das variáveis, em geral, não resultou em alta variação da APER. Apenas quando se retirou a variável “Duração da Gestação” houve uma diminuição da APER, ou seja, a classificação geral melhorou, porém a classificação correta das crianças que tiveram óbito diminuiu. Verificou-se também que a retirada da variável “Apgar no 5º Minuto” resultou no maior valor da APER, indicando que, possivelmente, a interação dessa variável com as demais é importante para a classificação correta dos indivíduos.

Usando as três variáveis consideradas pela literatura médica importantes na determinação de sobrevivência ou óbito no primeiro ano de vida, os resultados foram:

TABELA 4.4 – ANÁLISE DOS RESULTADOS UTILIZANDO AS TRÊS VARIÁVEIS CONSIDERADAS IMPORTANTES PELA LITERATURA MÉDICA

VARIÁVEIS UTILIZADAS	APER	PORCENTAGEM DE INDIVÍDUOS CORRETAMENTE CLASSIFICADOS	
		SOBREVIVÊNCIA	ÓBITO
Todas as variáveis	9,27 %	91,06 %	63,27 %
Apgar1, Apgar 5 e Peso	7,19 %	93,25 %	55,75 %
Idade da Mãe, Estado Civil da Mãe, Escolaridade da Mãe, Duração da Gestação, Tipo de Gravidez, Tipo de Parto, Número de Consultas de Pré-Natal, Sexo e Número de Gestações Anteriores	9,58 %	90,85 %	54,42 %

FONTE: O AUTOR

O uso das três variáveis consideradas importantes para a determinação da morte (ou

não) de uma criança no primeiro ano de vida apresentou o menor valor da APER quando comparada com o valor da APER obtida com as demais variáveis, ou seja, essas variáveis, sozinhas, classificam melhor a população geral. Porém, a classificação correta para a população com óbito foi prejudicada.

Usando as quatro variáveis consideradas pela literatura médica importantes na determinação de sobrevivência ou óbito no primeiro ano de vida, os resultados foram:

TABELA 4.5 – ANÁLISE DOS RESULTADOS UTILIZANDO AS QUATRO VARIÁVEIS CONSIDERADAS IMPORTANTES PELA LITERATURA MÉDICA

VARIÁVEIS UTILIZADAS	APER	PORCENTAGEM DE INDIVÍDUOS CORRETAMENTE CLASSIFICADOS	
		SOBREVIVÊNCIA	ÓBITO
Todas as variáveis	9,27 %	91,06 %	63,27 %
Idade da Mãe, Apgar1, Apgar 5 e Peso	7,29 %	93,16 %	55,75 %
Estado Civil da Mãe, Escolaridade da Mãe, Duração da Gestação, Tipo de Gravidez, Tipo de Parto, Número de Consultas de Pré-Natal, Sexo e Número de Gestações Anteriores	9,46 %	90,99 %	53,54 %

FONTE: O AUTOR

Quando se faz uso das quatro variáveis consideradas importantes para a determinação da morte ou não de uma criança no primeiro ano de vida também tem-se o menor valor da APER em comparação com o valor da APER obtido com as demais variáveis, ou seja, essas variáveis sozinhas também classificam melhor a população em geral.

Porém, novamente, a classificação correta para a população com óbito foi prejudicada. Com probabilidade a priori de óbito igual a 1,2% e de sobrevivência igual a 98,8% e excluindo as variáveis estudadas, uma a uma, e analisando o desempenho da função de reconhecimento e classificação, os resultados foram:

TABELA 4.6 – ANÁLISE DOS RESULTADOS EXCLUINDO AS VARIÁVEIS UMA A UMA

VARIÁVEL EXCLUÍDA	APER	PORCENTAGEM DE INDIVÍDUOS CORRETAMENTE CLASSIFICADOS	
		SOBREVIVÊNCIA	ÓBITO
Nenhuma	4,61 %	95,91 %	52,21 %
Idade da Mãe	4,60 %	95,92 %	51,77 %
Estado Civil da Mãe	4,59 %	95,93 %	51,77 %
Escolaridade da Mãe	4,62 %	95,91 %	51,33 %
Duração da Gestação	4,28 %	96,31 %	46,90 %
Tipo de Gravidez	3,30 %	97,26 %	50,00 %
Tipo de Parto	4,64 %	95,89 %	51,33 %
Número de Consultas de Pré-Natal	4,54 %	95,98 %	52,21 %
Sexo	4,64 %	95,88 %	51,77 %
Apgar no 1º Minuto	4,07 %	96,46 %	51,77 %
Apgar no 5º Minuto	4,44 %	96,09 %	51,33 %
Peso	4,60 %	95,94 %	50,00 %
Número de Gestações Anteriores	4,54 %	96,00 %	50,44 %

FONTE: O AUTOR

Observou-se que a retirada individual de cada uma das variáveis, em geral, não resultou em alta variação da APER. Apenas quando se retiraram as variáveis “Duração da Gestação” e “Tipo de Gravidez” houve uma diminuição da APER, ou seja, a classificação geral melhorou, porém a classificação correta das crianças que tiveram óbito diminuiu. A retirada da variável “Apgar no 1º Minuto” também resultou numa classificação melhor para a classificação geral.

Usando as três variáveis consideradas pela literatura médica importantes na determinação de sobrevivência ou óbito no primeiro ano de vida, os resultados foram:

TABELA 4.7 – ANÁLISE DOS RESULTADOS UTILIZANDO AS TRÊS VARIÁVEIS CONSIDERADAS IMPORTANTES PELA LITERATURA MÉDICA

VARIÁVEIS UTILIZADAS	APER	PORCENTAGEM DE INDIVÍDUOS CORRETAMENTE CLASSIFICADOS	
		SOBREVIVÊNCIA	ÓBITO
Todas as variáveis	4,61 %	95,91 %	52,21 %
Apgar1, Apgar 5 e Peso	2,80 %	97,85 %	42,92 %

VARIÁVEIS UTILIZADAS	APER	PORCENTAGEM DE INDIVÍDUOS CORRETAMENTE CLASSIFICADOS	
		SOBREVIVÊNCIA	ÓBITO
Idade da Mãe, Estado Civil da Mãe, Escolaridade da Mãe, Duração da Gestação, Tipo de Gravidez, Tipo de Parto, Número de Consultas de Pré-Natal, Sexo e Número de Gestações Anteriores	3,76 %	96,98 %	35,40 %

FONTE: O AUTOR

Por outro lado, o uso das três variáveis consideradas importantes para a determinação da morte (ou não) de uma criança no primeiro ano de vida apresentou o menor valor da APER quando comparada com o valor da APER obtido com as demais variáveis, ou seja, essas variáveis sozinhas classificam melhor a população geral. Porém, a classificação correta para a população com óbito foi prejudicada.

Usando as quatro variáveis consideradas pela literatura médica importantes na determinação de sobrevivência ou óbito no primeiro ano de vida:

TABELA 4.8 – ANÁLISE DOS RESULTADOS UTILIZANDO AS QUATRO VARIÁVEIS CONSIDERADAS IMPORTANTES PELA LITERATURA MÉDICA

VARIÁVEIS UTILIZADAS	APER	PORCENTAGEM DE INDIVÍDUOS CORRETAMENTE CLASSIFICADOS	
		SOBREVIVÊNCIA	ÓBITO
Todas as variáveis	4,61 %	95,91 %	52,21 %
Idade da Mãe, Apgar1, Apgar 5 e Peso	2,83 %	97,82 %	43,36 %
Estado Civil da Mãe, Escolaridade da Mãe, Duração da Gestação, Tipo de Gravidez, Tipo de Parto, Número de Consultas de Pré-Natal, Sexo e Número de Gestações Anteriores	3,76 %	96,98 %	35,40 %

FONTE: O AUTOR

Finalmente, observou-se que o uso das quatro variáveis consideradas importantes para a determinação da morte (ou não) de uma criança no primeiro ano de vida também apresentou o menor valor da APER quando comparada com o valor da APER obtido com as demais variáveis, ou seja, essas variáveis, sozinhas, também classificam melhor a população geral. Porém, a classificação correta para a população com óbito foi prejudicada. A inclusão da variável “Idade da Mãe” não resultou em diferenças significativas na classificação.

CONCLUSÕES

A função de reconhecimento e classificação respondeu de maneira satisfatória para a classificação geral dos indivíduos, principalmente considerando a probabilidade a priori de óbito igual a 1,2%. No melhor resultado, obtido com as três variáveis (Apgar 1, Apgar 5 e peso do recém-nascido) consideradas importantes pela literatura médica na determinação ou não do óbito, classificou corretamente 97,20% das observações. Porém esse desempenho é fortemente afetado pelo grupo sobrevivente ao primeiro ano de vida que corresponde a 98,8% dos dados analisados. A classificação correta para as crianças com óbito no primeiro ano de vida, neste caso, ficou em 42,92%. Usando-se as probabilidades a priori iguais a 50%, o melhor resultado obtido classificou corretamente 92,81% das observações, também usando as variáveis Apgar 1, Apgar 5 e peso do recém-nascido. Como o objetivo é classificar corretamente, principalmente, as crianças com probabilidade de óbito no primeiro ano de vida, o melhor resultado foi a classificação correta de 64,60% das crianças que chegaram a óbito, obtida com a exclusão da variável “Tipo de Gravidez” e probabilidades a priori de óbito e de sobrevivência iguais. Neste caso, a classificação correta para as crianças com sobrevivência no primeiro ano de vida ficou em 91,09%. Sem a exclusão de variáveis, o melhor resultado para a classificação correta das crianças que chegaram a óbito foi 63,27%, com 91,06% de classificação correta para as crianças sobreviventes, novamente com probabilidades a priori iguais a 50%.

Em todos os testes, as variáveis indicadas pela literatura médica mostraram que, realmente, são mais discriminantes do que as demais variáveis analisadas para a classificação geral (dos dois grupos: sobreviventes e não-sobreviventes). Porém, para uma melhor classificação geral e que atenda, também, a uma melhor classificação das crianças com probabilidade de óbito recomenda-se que nenhuma das variáveis estudadas seja excluída no momento de construir a função de reconhecimento e classificação e na realização da classificação de um novo indivíduo. Para melhores resultados nesse sentido, deve-se usar, também, probabilidades a priori iguais.

Recomenda-se que os registros das declarações de óbitos sejam codificados de

forma conjugada às declarações de nascidos vivos para integrar as bases de dados dos dois sistemas, SIM e SINASC, e assim, facilitar o cruzamento dos dados das duas bases com conseqüente ganho de informações que, se agregadas, permitem estudos com maior confiabilidade.

Espera-se que este estudo possa somar juntamente com outros realizados na área quanto a medidas de prevenção da mortalidade infantil, alertando para as variáveis com alto grau de associação na ocorrência de óbito, e assim, monitorar os recém-nascidos que devem receber assistência efetiva e servindo de auxílio para trabalhos futuros.

O modelo desenvolvido para reconhecimento e classificação permite implantar um sistema de vigilância, em hospitais ou em outros estabelecimentos onde ocorra partos, serviços de vigilância epidemiológica e nas regionais de saúde, para auxiliar na tomada de decisões e prevenir óbitos no primeiro ano de vida.

REFERÊNCIAS

- ABC DA SAÚDE. **Índice de Apgar**. Disponível em: <http://www.abcdasaude.com.br/artigo.php?254>, Data de acesso: 15/06/2006
- ALMEIDA, M.O. **Uma ferramenta para mineração visual de dados usando mapas em árvore e suas aplicações**. Salvador: Universidade Salvador, 2003. Disponível em: http://www.energia.unifacs.br/dissertacoes/marcio_almeida.pdf, Data de acesso: 20/03/2006.
- ANDRADE, S.M.; SOARES, D.A.; MATSUO, T.; SOUZA, R.K.T.; MATHIAS, T.A.F.; IWAKUARA, M.L.H.; ZEQUIM, M.A. **Condições de vida e mortalidade infantil no Estado do Paraná, Brasil, 1997/2001**. Disponível em: <http://www.scielo.org/pdf/csp/v22n1/19.pdf>, Data de acesso: 15/06/2006.
- BARSA. **Estatística**. Enciclopaedia Britannica do Brasil, 2001 [CD-ROM]
- CALDEIRA, A.P.; FRANÇA, E.; GOULART, E.M.A. **Mortalidade infantil pós-neonatal e qualidade da assistência médica: um estudo caso-controle**. J. Pediatr. (Rio de J.) vol.77 no.6 Porto Alegre Nov./Dec. 2001. Disponível em: <http://www.scielo.br/pdf/jped/v77n6/v77n6a08.pdf>. Data de acesso: 15/05/2006.
- CARVALHO, F. P.; FAGUNDES JR., A.; SILVEIRA, J.G.; AZAMBUJA, M.C. **KDD-NMS: Um sistema de descoberta de conhecimento e mineração em bases de dados de sistemas de gerência de redes**. Disponível em: www.rnp.br/_arquivo/wrnp2/2003/oscfa01a.pdf, Data de acesso: 20/03/2006.
- CARREIRA, M.R. **Critérios para classificação visual de peças estruturais de *pinus sp.*** Disponível em: <http://www.teses.usp.br/teses/disponiveis/18/18134/tde-02082003-215557/publico/mrcarreira05.pdf>. Data de acesso: 15/06/2006
- CHAVES NETO, A. **Análise multivariada II: notas de aula**. Curitiba: UFPR, 2005.
- COSTA, S.C. **Regressão logística aplicada na identificação de fatores de risco para doenças em animais domésticos**. Dissertação de Mestrado. Piracicaba: Escola Superior de Agricultura “Luiz de Queiroz”, 1997.
- CRUZ, M.C.C. **O impacto da amamentação sobre a desnutrição e a mortalidade infantil, Brasil, 1996**. [Mestrado] Fundação Oswaldo Cruz, Escola Nacional de Saúde Pública; 2001. 80 p. Disponível em: <http://portalteses.cict.fiocruz.br/>, Data de acesso: 20/03/2006.
- DEMÉTRIO, C.G.B. **Modelos lineares generalizados na experimentação agrônômica**. In: Reunião anual da região brasileira da sociedade internacional de biometria, 38.; simpósio de estatística aplicada à experimentação agrônômica, 5., 1993, Porto Alegre. Anais... Porto Alegre: UFRS, 1993. p.127.
- GUIMARÃES, I.A. **Construção e avaliação de uma regra de reconhecimento e classificação de clientes de uma instituição financeira com base na análise multivariada**. Dissertação de Mestrado. Curitiba: UFPR, 2000.
- GUIMARÃES, I.A.; CHAVES NETO, A. **Reconhecimento de padrões: metodologias estatísticas em crédito ao consumidor**. ERA-eletronica, vol.1 n.2, jul-dez/2002. Disponível em: <http://www.rae.com.br/electronica>. Data de acesso: 20/03/2006.
- _____. **Reconhecimento de padrões: comparação de métodos multivariados e redes neurais**. Disponível em: rnti.fesppr.br/include/getdoc.php?id=92&article=4&mode=pdf,

Data de acesso: 21/05/2006

HAIR, J.F.; ANDERSON, R. E.; TATHAM, R. L.; BLACK, W. C.. **Análise multivariada de dados**. 5. ed. Porto Alegre: Bookman, 2005.

IBGE. **A mortalidade no Brasil no período 1980 – 2004: desafios e oportunidades para os próximos anos**. Disponível em: <http://www.ibge.gov.br>. Data de acesso: 20/06/2006.

_____. **Cidades**. Disponível em: <http://www.ibge.gov.br>. Data de acesso: 20/06/2006.

JBONLINE. **Impacto na taxa de mortalidade infantil do país**. 2004. Disponível em: <http://jbonline.terra.com.br/jb/papel/economia/2004/05/29>. Data de acesso: 20/06/2006

JOHNSON, R.A.; WICHERN, D.W. **Applied multivariate statistical analysis**. 4. ed. New Jersey: Prentice-Hall, 1998.

KOZU, K.; GODINHO, L.T.; MUNIZ ; M.V.F.; CHIARIONI, P. **Mortalidade infantil: causas e fatores de risco. um estudo bibliográfico**. Disponível em: <http://www.medstudents.com.br/original/original/mortinf/mortinf.htm>. Data de acesso: 02/03/2006.

LIMA, J.D. **A análise econômico-financeira de empresas sob a ótica da estatística multivariada**. Dissertação de Mestrado. Curitiba: UFPR, 2002.

LOPES, P.A. **Entendendo a importância da estatística sem ser gênio, matemático ou bruxo**. Disponível em: http://www.sindhosp.com.br/IMP_291D.htm, Data de acesso: 21/05/2006.

MOITA NETO, J. M. **Estatística multivariada: uma visão didática-metodológica**. Disponível em: http://www.criticanarede.com/cien_estadistica.html, Data de acesso: 21/05/2006.

MAGRO FILHO, J.B.; RIBEIRO, M.A.A. **Mortalidade infantil**. Disponível em: <http://www.aborto.com.br/mortalidade/index.htm>. Data de acesso: 02/03/2006

MARQUES, M.A.M. **Aplicação da análise multivariada no estudo da infra-estrutura dos serviços de saúde dos municípios paranaenses**. Dissertação de Mestrado. Curitiba: UFPR, 2006.

MINISTÉRIO DA SAÚDE. **Sistema de informações sobre nascidos vivos – SINASC**. Disponível em: http://dtr2001.saude.gov.br/svs/sis/sis00_SINASC.htm. Data de acesso: 02/03/2006.

MISHIMA, F.C. **Declaração de nascido vivo: análise do seu preenchimento no Município de Ribeirão Preto, São Paulo, Brasil**. Cad. Saúde Pública v.15 n.2 Rio de Janeiro abr./jun. 1999. Disponível em: <http://www.scielo.br/pdf/csp/v15n2/0323.pdf>. Data de acesso: 15/05/2006.

MÜLLER, L. **Uma loja virtual adaptativa com base no comportamento navegacional e compras**. Passo Fundo: Universidade de Passo Fundo, 2002. Disponível em: www.upf.br/computacao/download/tcii_luciane.pdf, Data de acesso: 20/03/2006.

NOVAES, H.M.D.; ALMEIDA, M.F. ; ORTIZ, L.P. **Gestão para redução da mortalidade infantil**. Disponível em: <http://itd.bvs.br/>. Data de acesso: 20/02/2006.

OLIVEIRA, I.C. **Aplicação de data mining em busca de um modelo de prevenção da mortalidade infantil**. Dissertação de Mestrado. Florianópolis: UFSC, 2001.

QUISPER, N. R. P. **Técnicas e ferramentas para a extração inteligente e automática de conhecimento em banco de dados**. Campinas: Universidade Estadual de Campinas, 2003. Disponível em: www.libdigi.unicamp.br/document/?code=vtls000308439, Data de acesso: 20/03/2006.

RAMOS, E.M.L.S. **Estatística: poderosa ciência ao alcance de todos**. Disponível em: <http://www.ufpa.br/beiradorio/arquivo/Beira21/opiniaio.html>, Data de acesso: 21/05/2006.

SANTA HELENA, E.T. **Fatores de risco para natalidade neonatal em Blumenau, Santa Catarina: linkage entre banco de dados**. Rev. Brasileira de Saúde Infantil. vol. 5, n.2, Recife: 2005.

SANTOS, O.J.S.; MILIONI, A.Z. **Composição de especialistas locais para classificação de dados**. Disponível em: <http://www.scielo.oces.mctes.pt/pdf/iop/v25n1/v25n1a06.pdf>. Data de acesso: 15/05/2006.

SCARPEL, R.A. **Modelos de previsão de insolvência: uma abordagem discriminante paramétrica e não paramétrica**. São José dos Campos: ITA/IMB, 2005


SOUZA, R.K.T.; GOTLIEB, S.L.D. **Probabilidade de morrer no primeiro ano de vida em área urbana da região sul, Brasil**. Re. Saúde Pública. São Paulo: FAPESP, n.27, 1993, p. 445-54.

TOMÉ, E.A.; LATORRE, M.R.D.O. **Tendências da mortalidade infantil no Município de Guarulhos: análise do período de 1971 a 1998**. Disponível em: <http://www.scielo.br/pdf/rbepid/v4n3/03.pdf> Data de acesso: 15/05/2006

VON WANGENHEIM, A. **Reconhecimento de padrões: técnicas estatísticas**. Disponível em: <http://www.inf.ufsc.br/~awangenh/RP/estatisticas.html>, Data de acesso: 21/04/2006.

ZANDONADE, E. **Estatística judiciária, importância e meios**. Disponível em: <http://www.cjf.gov.br/revista/numero17/painel33.pdf>, Data de acesso: 21/05/2006.

ANEXO I – DECLARAÇÃO DE NASCIDO VIVO (DN)

 República Federativa do Brasil Ministério da Saúde 1ª Via - Secretaria de Saúde		Declaração de Nascido Vivo Nº _____	
I	Cartório	1	Cartório
		2	Registro
II	Local da Ocorrência	3	Data
		4	Município
		5	UF
		6	Local da ocorrência
III	Mãe	7	Estabelecimento
		8	Endereço da ocorrência, se fora do estab. ou da resid. da mãe (Rua, praça, avenida, etc)
		9	CEP
		10	Bairro/Distrito
IV	Gestação e parto	11	Município de ocorrência
		12	UF
		13	Nome da mãe
		14	RIC
V	Recém Nascido	15	Idade (anos)
		16	Estado civil
		17	Escolaridade (Emonce de estudo concluída)
		18	Ocupação habitual e ramo de atividade
VI	Identificação	19	Núm de filhos tidos em gestações anteriores (obs: utilizar 99 se ignorados)
		20	Logradouro
		21	CEP
		22	Bairro/distrito
VII	Atenção	23	Duração da gestação (em semanas)
		24	Tipo de gravidez
		25	Tipo de parto
		26	Número de consultas de pré-natal
VIII	Resumo	27	Nascimento
		28	Raça/cor
		29	Sexo
		30	Peso ao nascer
IX	Observações	31	Índice de Apgar
		32	Detectada alguma malformação congênita e/ou anomalia cromossômica?
		33	Polegar direito da mãe
		34	Pé direito da criança
X	Responsável pelo preenchimento	35	Nome
		36	Função
		37	Identidade
		38	Órgão Emissor
XI	Data	39	Data
		40	Data
		41	Data
		42	Data

ATENÇÃO : ESTE DOCUMENTO NÃO SUBSTITUI A CERTIDÃO DE NASCIMENTO

O Registro de Nascimento é obrigatório por lei.

Para registrar esta criança, o pai ou responsável deverá levar este documento ao cartório de registro civil.

Versão 0008-1

ANEXO II – DECLARAÇÃO DE ÓBITO (DO)

República Federativa do Brasil
Ministério da Saúde
1ª VIA - SECRETARIA DE SAÚDE

Declaração de Óbito

I	Cidade	1) Cód. do Município	2) Registro	3) Data
		4) Município	5) UF	6) Cemitério
II	Identificação	7) Tipo de Óbito	8) Cód. do Local	9) RIC
		10) Nome do falecido	11) Nome da mãe	12) Naturalidade
III	Residência	13) Data de nascimento	14) Idade	15) Sexo
		16) Estado civil	17) Escolaridade	18) Ocupação habitual e ramo de atividade
IV	Ocorrência	19) Logradouro (Rua, praça, avenida etc.)	20) Cód. do Município de residência	21) CEP
		22) Bairro/Distrito	23) Cód. do Município de ocorrência	24) CEP
V	Fetal ou menor que 1 ano	25) Local de ocorrência do óbito	26) Estabelecimento	27) Cód. do Município de ocorrência
		28) Endereço de ocorrência, se fora do estabelecimento ou da residência (Rua, praça, avenida etc.)	29) Cód. do Município de ocorrência	30) CEP
VI	Condições e causas do óbito	PREENCHIMENTO EXCLUSIVO PARA ÓBITOS FETAIS E DE MENORES DE 1 ANO		
		INFORMAÇÕES SOBRE A MÃE		
VII	Médico	CAUSAS DA MORTE		
		CAUSAS ANTERIORES		
VIII	Causas externas	PROVÁVEIS CIRCUNSTÂNCIAS DE MORTE NÃO NATURAL		
		SE A OCORRÊNCIA FOR EM VIA PÚBLICA, ANOTAR O ENDEREÇO		
IX	Localidade	DECLARAÇÃO		
		TOSCOMUNHAS		

ANEXO III – CÓDIGO FONTE DO PROGRAMA

```

unit UniPri;

interface

uses
  Windows, Messages, SysUtils, Variants, Classes, Graphics,
  Controls, Forms,
  Dialogs, ZConnection, StdCtrls, Grids, DBGrids, DB,
  ZAbstractRODataset,
  ZAbstractDataset, ZAbstractTable, ZDataset, ExtCtrls, DBCtrls,
  ComCtrls,
  AppEvnts, Gauges, Spin, CheckLst;

type
  TfrmPri = class(TForm)
    cnx: TZConnection;
    TabPes: TZTable;
    DataSource1: TDataSource;
    OpenDialog1: TOpenDialog;
    TabPesCODPES: TIntegerField;
    TabPesC01: TFloatField;
    TabPesC02: TFloatField;
    TabPesC03: TFloatField;
    TabPesC04: TFloatField;
    TabPesC05: TFloatField;
    TabPesC06: TFloatField;
    TabPesC07: TFloatField;
    TabPesC08: TFloatField;
    TabPesC09: TFloatField;
    TabPesC10: TFloatField;
    TabPesC11: TFloatField;
    TabPesC12: TFloatField;
    TabPesC13: TFloatField;
    TabPesC14: TFloatField;
    TabPesC15: TFloatField;
    TabPesGRUPES: TStringField;
    Que: TZReadOnlyQuery;
    ApplicationEvents1: TApplicationEvents;
    DataSource2: TDataSource;
    PageControl1: TPageControl;
    TabSheet1: TTabSheet;
    TabSheet2: TTabSheet;
    DBGrid1: TDBGrid;
    GroupBox1: TGroupBox;
    ComboBox1: TComboBox;
    Button1: TButton;
    Button4: TButton;
    GroupBox2: TGroupBox;
    Label1: TLabel;
    tarefas: TCheckListBox;
    Button3: TButton;
    priori: TEdit;
  end;

```

```

    Button2: TButton;
    resultado: TStringGrid;
    Button6: TButton;
    GroupBox3: TGroupBox;
    Label2: TLabel;
    Novo: TStringGrid;
    Button5: TButton;
    resind: TEdit;
    ListBox1: TListBox;
    tx: TEdit;
    Label3: TLabel;
    procedure Button1Click(Sender: TObject);
    procedure FormCreate(Sender: TObject);
    procedure FormClose(Sender: TObject; var Action: TCloseAction);
    procedure Button3Click(Sender: TObject);
    procedure TabPesAfterOpen(DataSet: TDataSet);
    procedure ApplicationEvents1Exception(Sender: TObject; E:
Exception);
    procedure Button4Click(Sender: TObject);
    procedure Button2Click(Sender: TObject);
    procedure Button5Click(Sender: TObject);
    procedure Button6Click(Sender: TObject);
private
    { Private declarations }
public
    { Public declarations }
end;

var
    frmmpri: Tfrmmpri;

    i,j,k,l : integer;                //contadores
    numvar : integer;                 //número de
variáveis
    covarv,covarm : array[1..16,1..16] of real;    //matrizes de
covariancia
    vetmdv,vetmdm : array[1..16] of real;          //vetor medios
    quantv,quantm : integer;                       //quantidades de
vivos e mortos
    som1,som2,som3 : real;                          //acumulador
    inverv,inverm : array[1..16,1..32] of real;      //matrizes
inversas ampliadas
    pivo : array[1..16,1..16] of real;               //pivo para
calculo das inversas
    cholev,cholem : array[1..16,1..16] of real;      //matrizes
decompostas de cholesky
    deterv,determ : real;                          //determinantes
    somav,somam : real;                            //soma de vivos e
mortos
    contvn,contvp,contmn,contmp:integer;            //contadores de
valores negativos e positivos
    res1,res2,res3 : array[1..16] of real;           //resultados dos
produtos de vetores por matrizes
    coefk: real;                                    //coeficiente K
    difinv : array[1..16,1..16] of real;             //matriz

```

```

diferença de inversa
    prioriv,priorim : real;                                //probabilidades
a priori

implementation

{$R *.dfm}

procedure TfrmPri.Button1Click(Sender: TObject);
var f : textfile;
    linha : string;
    c: real;
    i,campo,cont: integer;
begin
    listBox1.Clear;
    listBox1.Items.Add(timetostr(time)+' Importando dados');
    try
        dbgrid1.DataSource:=datasource2;
        if opendialog1.Execute then begin
            assignfile(f,opendialog1.FileName);
            reset(f);
            cont:=0;
            while not eof(f) do begin
                readln(f,linha);
                if copy(linha,length(linha)-1,1)<>';' then
                    linha:=linha+';';
                if copy(linha,1,2)=';' then exit;
                tabpes.Append;
                tabpesgrupes.Value:=copy(combobox1.Text,1,1);
                for campo:=2 to 17 do begin
                    //val(copy(linha,1,pos(';',linha)-1),c,i);
                    if trim(copy(linha,1,pos(';',linha)-1))<>' then
                        begin
                            c:=strtofloat(copy(linha,1,pos(';',linha)-1));
                            tabpes.Fields[campo].Value:=c;
                            delete(linha,1,pos(';',linha));
                        end;
                    end;
                tabpes.Post;
                cont:=cont+1;
                listBox1.Items.Strings[1]:='-->Progresso:
'+inttostr(cont);
                listBox1.Refresh;
            end;
            closefile(f);
        end;
        tabpes.refresh;
        {verificando quantidade de variáveis}
        for numvar:=0 to 16 do if tabpes.Fields[numvar+2].value=null
then break;
        finally
            dbgrid1.DataSource:=datasource1;
        end;
    end;
end;

```

```

procedure TfrmPri.FormCreate(Sender: TObject);
var lis:tstringlist;
    linha : string;
begin
    cnx.Connect;
    tabpes.active:=true;
    novo.Rows[0].Strings[0]:='Variáveis';
    novo.Rows[1].Strings[0]:='Valores';
    novo.Rows[0].Strings[1]:='Idade';
    novo.Rows[0].Strings[2]:='Est.Civil';
    novo.Rows[0].Strings[3]:='Escolaridade';
    novo.Rows[0].Strings[4]:='Gestação';
    novo.Rows[0].Strings[5]:='Gravidez';
    novo.Rows[0].Strings[6]:='Parto';
    novo.Rows[0].Strings[7]:='Consultas';
    novo.Rows[0].Strings[8]:='Sexo';
    novo.Rows[0].Strings[9]:='Apgar1';
    novo.Rows[0].Strings[10]:='Apgar5';
    novo.Rows[0].Strings[11]:='Peso';
    novo.Rows[0].Strings[12]:='NºGestações';
    resultado.Rows[0].Strings[1]:='Vivos';
    resultado.Rows[0].Strings[2]:='Óbitos';
    resultado.Rows[1].Strings[0]:='Dos Vivos';
    resultado.Rows[2].Strings[0]:='Dos Mortos';
    for numvar:=0 to 16 do if tabpes.Fields[numvar+2].value=null then
break;
    //////////////////////////////////ler classificação////////////////////////////////////
    listbox1.Items.Add(timetostr(time)+' Lendo classificação');
    lis:=tstringlist.create;
    try
        if fileexists('vars.ini') then begin
            lis.LoadFromFile('vars.ini');
        end else exit;
        priori.text:=copy(lis.Strings[0],9,255);
        coefk:=strtofloat(copy(lis.Strings[1],9,255));
        linha:=copy(lis.Strings[2],9,length(lis.Strings[2]));
        for k:=1 to numvar do begin
            vetmdv[k]:=strtofloat(copy(linha,1,pos(';',linha)-1));
            delete(linha,1,pos(';',linha));
        end;
        linha:=copy(lis.Strings[3],9,length(lis.Strings[3]));
        for k:=1 to numvar do begin
            vetmdm[k]:=strtofloat(copy(linha,1,pos(';',linha)-1));
            delete(linha,1,pos(';',linha));
        end;
        for j:=1 to numvar do begin
            linha:=copy(lis.Strings[j+3],9,length(lis.Strings[j+3]));
            for k:=1 to numvar do begin
                difinv[j,k]:=strtofloat(copy(linha,1,pos(';',linha)-1));
                delete(linha,1,pos(';',linha));
            end;
        end;
        for j:=1 to numvar do begin
            linha:=copy(lis.Strings[j+numvar+3],9,length(lis.Strings[j+numvar+3])

```

```

));
    for k:=1 to numvar do begin
        inverv[j,k]:=strtofloat(copy(linha,1,pos('; ',linha)-1));
        delete(linha,1,pos('; ',linha));
    end;
end;
for j:=1 to numvar do begin
    linha:=copy(lis.Strings[j+numvar*2+3],9,length(lis.Strings[j+numvar*
2+3]));
    for k:=1 to numvar do begin
        invern[j,k]:=strtofloat(copy(linha,1,pos('; ',linha)-1));
        delete(linha,1,pos('; ',linha));
    end;
end;
finally
    lis.free;
end;
end;

procedure Tfrmprj.FormClose(Sender: TObject; var Action:
TCloseAction);
begin
    cnx.Disconnect;
    application.Terminate;
end;

procedure Tfrmprj.Button3Click(Sender: TObject);
begin
    for i:=0 to tarefas.Count-1 do begin
        tarefas.Checked[i]:=false;
    end;
    try
        dbgrid1.DataSource:=datasource2;
        listbox1.Items.Add(timetoepr(time)+' Calculando vetor médio');
        listbox1.Repaint;
        {calculando vetor médio}
        que.SQL.Clear;
        que.SQL.add('select count(codpes) as c from origem where
grupes='1'');
        que.Active:=true;
        quantv:=que.FieldList.Fields[0].AsInteger;
        listbox1.Items.Add(timetoepr(time)+' -->População vivos :
'+inttostr(quantv));
        listbox1.Repaint;
        que.SQL.Clear;
        que.SQL.add('select count(codpes) as c from origem where
grupes='2'');
        que.Active:=true;
        quantm:=que.FieldList.Fields[0].AsInteger;
        listbox1.Items.Add(timetoepr(time)+' -->População mortos :
'+inttostr(quantm));
        listbox1.Repaint;
        que.Active:=false;
        que.SQL.Clear;
    end;
end;

```

```

        que.SQL.add('select avg(c01) as mc01, avg(c02) as mc02,
avg(c03) as mc03, avg(c04) as mc04, avg(c05) as mc05,');
        que.SQL.add(' avg(c06) as mc06, avg(c07) as mc07, avg(c08) as
mc08, avg(c09) as mc09, avg(c10) as mc10,');
        que.SQL.Add(' avg(c11) as mc11, avg(c12) as mc12, avg(c13) as
mc13, avg(c14) as mc14, avg(c15) as mc15 from ORIGEM where
grupes='1');
        que.Active:=true;
        for i:=1 to 15 do vetmdv[i]:=que.FieldList.Fields[i-
1].AsFloat;
        que.Active:=false;
        que.SQL.Clear;
        que.SQL.add('select avg(c01) as mc01, avg(c02) as mc02,
avg(c03) as mc03, avg(c04) as mc04, avg(c05) as mc05,');
        que.SQL.add(' avg(c06) as mc06, avg(c07) as mc07, avg(c08) as
mc08, avg(c09) as mc09, avg(c10) as mc10,');
        que.SQL.Add(' avg(c11) as mc11, avg(c12) as mc12, avg(c13) as
mc13, avg(c14) as mc14, avg(c15) as mc15 from ORIGEM where
grupes='2');
        que.Active:=true;
        for i:=1 to 15 do vetmdm[i]:=que.FieldList.Fields[i-
1].AsFloat;
        que.Active:=false;
        listbox1.Items.Add(timetostr(time)+' Calculando covariância
para vivos');
        listbox1.Repaint;
        {calculando covariancia de vivos}
        for i:=1 to numvar do for j:=1 to numvar do covarv[i,j]:=0;
        ///// diagonal /////
        tabpes.Filter:='grupes='1';
        tabpes.Filtered:=true;
        for l:=1 to numvar do begin
            if tabpes.Fields[l+1].value=null then break;
            tabpes.First; //vai para a primeira pessoa
            soml:=0; //acumulador
            while not tabpes.eof do begin
                if tabpes.grupes.Value='1' then
                    soml:=soml+(tabpes.Fields[l+1].Value-
vetmdv[l])*(tabpes.Fields[l+1].Value-vetmdv[l]);
                tabpes.next; //proximo registro de pessoas
            end;
            if soml>0 then covarv[l,l]:=soml/(quantv-1);
        end;
        ///////////////////////////////////
        for i:=1 to numvar do begin
            for j:=i+1 to numvar do begin
                if tabpes.fields[j+1].value=null then break;
                soml:=0;
                tabpes.First;
                while not tabpes.eof do begin
                    if tabpes.grupes.Value='1' then
                        soml:=soml+(tabpes.Fields[i+1].Value-
vetmdv[i])*(tabpes.Fields[j+1].value-vetmdv[j]);
                    tabpes.next;
                end;
            end;
        end;
    end;

```

```

        soml:=soml/(quantv-1);
        covarv[j,i]:=soml;
        covarv[i,j]:=soml;
    end;
end;
tarefas.Checked[0]:=true;
listbox1.Items.Add(timetostr(time)+'    Calculando    covariância
para mortos');
listbox1.Repaint;
{calculando covariancia de mortos}
for i:=1 to numvar do for j:=1 to numvar do covarm[i,j]:=0;
///// diagonal /////
tabpes.Filter:='grupes=''2''';
tabpes.Filtered:=true;
for l:=1 to numvar do begin
    if tabpes.Fields[l+1].value=null then break;
    tabpes.First; //vai para a primeira pessoa
    soml:=0;      //acumulador
    while not tabpes.eof do begin
        if
            tabpesgrupes.Value='2'
        then
            soml:=soml+(tabpes.Fields[l+1].Value-
vetmdm[l])*(tabpes.Fields[l+1].Value-vetmdm[l]);
            tabpes.next; //proximo registro de pessoas
        end;
        if soml>0 then covarm[l,l]:=soml/(quantm-1);
    end;
    ///////////////////////////////////
for i:=1 to numvar do begin
    for j:=i+1 to numvar do begin
        if tabpes.fields[j+1].value=null then break;
        soml:=0;
        tabpes.First;
        while not tabpes.eof do begin
            if
                tabpesgrupes.Value='2'
            then
                soml:=soml+(tabpes.Fields[i+1].Value-
vetmdm[i])*(tabpes.Fields[j+1].value-vetmdm[j]);
                tabpes.next;
            end;
            soml:=soml/(quantm-1);
            covarm[j,i]:=soml;
            covarm[i,j]:=soml;
        end;
    end;
end;
tabpes.filtered:=false;
tarefas.Checked[1]:=true;
listbox1.Items.Add(timetostr(time)+'    Calculando inversão para
vivos');
listbox1.Repaint;
for k:=1 to numvar do begin //linha
    for i:=1 to numvar do begin //coluna
        inverv[k,i]:=covarv[k,i];
        if
            k=i
        then
            inverv[k,i+numvar]:=1
        else
            inverv[k,i+numvar]:=0;
        end;
    end;
end;

```

```

////////////////////////////////////
for k:=1 to numvar-1 do begin
  for i:=k+1 to numvar do begin
    if inverv[k,k]=0 then begin
      for l:=1 to numvar*2 do begin
        soml:=inverv[i,l];
        inverv[i,l]:=inverv[i+1,l];
        inverv[i+1,l]:=soml;
      end;
      end else pivo[i,k]:=inverv[i,k]/inverv[k,k];
      for j:=k to numvar*2 do begin
        inverv[i,j]:=inverv[i,j]-pivo[i,k]*inverv[k,j];
      end;
    end;
  end;
end;
for k:=1 to numvar-1 do begin
  for i:=k+1 to numvar do begin
    if inverv[i,i]=0 then begin
      for l:=1 to numvar*2 do begin
        soml:=inverv[i,l];
        inverv[i,l]:=inverv[i+1,l];
        inverv[i+1,l]:=soml;
      end;
      end else pivo[k,i]:=inverv[k,i]/inverv[i,i];
      for j:=1 to numvar*2 do begin
        inverv[k,j]:=inverv[k,j]-pivo[k,i]*inverv[i,j];
      end;
    end;
  end;
end;
for i:=1 to numvar do begin
  pivo[i,i]:=1/inverv[i,i];
  for j:=i to numvar*2 do begin
    inverv[i,j]:=pivo[i,i]*inverv[i,j];
  end;
end;
////////////////////////////////////
tarefas.Checked[2]:=true;
listbox1.Items.Add(timetostr(time)+' Calculando inversão para
mortos');
listbox1.Repaint;
for k:=1 to numvar do begin //linha
  for i:=1 to numvar do begin //coluna
    inverm[k,i]:=covarm[k,i];
    if k=i then inverm[k,i+numvar]:=1 else
inverm[k,i+numvar]:=0;
  end;
end;
////////////////////////////////////
for k:=1 to numvar-1 do begin
  for i:=k+1 to numvar do begin
    if inverm[k,k]=0 then begin
      for l:=1 to numvar*2 do begin
        soml:=inverm[i,l];
        inverm[i,l]:=inverm[i+1,l];
        inverm[i+1,l]:=soml;

```



```

        end;
        end else pivo[i,k]:=inverm[i,k]/inverm[k,k];
        for j:=k to numvar*2 do begin
            inverm[i,j]:=inverm[i,j]-pivo[i,k]*inverm[k,j];
        end;
    end;
end;
for k:=1 to numvar-1 do begin
    for i:=k+1 to numvar do begin
        if inverm[i,i]=0 then begin
            for l:=1 to numvar*2 do begin
                soml:=inverm[i,l];
                inverm[i,l]:=inverm[i+1,l];
                inverm[i+1,l]:=soml;
            end;
            end else pivo[k,i]:=inverm[k,i]/inverm[i,i];
            for j:=1 to numvar*2 do begin
                inverm[k,j]:=inverm[k,j]-pivo[k,i]*inverm[i,j];
            end;
        end;
    end;
end;
for i:=1 to numvar do begin
    pivo[i,i]:=1/inverm[i,i];
    for j:=i to numvar*2 do begin
        inverm[i,j]:=pivo[i,i]*inverm[i,j];
    end;
end;
//////////////////////////////////////
tarefas.Checked[3]:=true;
{calculando decomposição de cholesky para vivos}
listbox1.Items.Add(timetostr(time)+' Calculando decomposição
de cholesky para vivos');
listbox1.Repaint;
cholev[1,1]:=sqrt(covarv[1,1]);
for j:=2 to numvar do begin
    cholev[j,1]:=covarv[j,1]/cholev[1,1];
end;
for i:=2 to numvar-1 do begin
    soml:=0;
    for k:=1 to i-1 do begin
        soml:=soml+cholev[i,k]*cholev[i,k];
    end;
    if covarv[i,i]-soml<0 then begin
        listbox1.Items.add(timetostr(time)+' -->A matriz não é
simétrica ou não é positiva definida...');
        exit;
    end else cholev[i,i]:=sqrt(covarv[i,i]-soml);
    for j:=i+1 to numvar do begin
        som2:=0;
        for k:=1 to i-1 do begin
            som2:=som2+cholev[j,k]*cholev[i,k];
        end;
        cholev[j,i]:=(covarv[j,i]-som2)/cholev[i,i];
    end;
end;
end;

```

```

som3:=0;
for k:=1 to numvar-1 do begin
    som3:=som3+cholev[numvar,k]*cholev[numvar,k];
end;
if covarv[numvar,numvar]-som3<0 then begin
    listbox1.Items.add(timetostr(time)+' -->A matriz não é
positiva definida...');
    exit;
end else cholev[numvar,numvar]:=sqrt(covarv[numvar,numvar]-
som3);
for i:=1 to numvar do begin
    for j:=2 to numvar do begin
        if j>i then cholev[i,j]:=0;
    end;
end;
////////////////////////////////////
tarefas.Checked[4]:=true;
{calculando decomposição de cholesky para mortos}
listbox1.Items.Add(timetostr(time)+' Calculando decomposição
de cholesky para mortos');
listbox1.Repaint;
cholem[1,1]:=sqrt(covarm[1,1]);
for j:=2 to numvar do begin
    cholem[j,1]:=covarm[j,1]/cholem[1,1];
end;
for i:=2 to numvar-1 do begin
    som1:=0;
    for k:=1 to i-1 do begin
        som1:=som1+cholem[i,k]*cholem[i,k];
    end;
    if covarm[i,i]-som1<0 then begin
        listbox1.Items.add(timetostr(time)+' -->A matriz não é
simétrica ou não é positiva definida...');
        exit;
    end else cholem[i,i]:=sqrt(covarm[i,i]-som1);
    for j:=i+1 to numvar do begin
        som2:=0;
        for k:=1 to i-1 do begin
            som2:=som2+cholem[j,k]*cholem[i,k];
        end;
        cholem[j,i]:=(covarm[j,i]-som2)/cholem[i,i];
    end;
end;
som3:=0;
for k:=1 to numvar-1 do begin
    som3:=som3+cholem[numvar,k]*cholem[numvar,k];
end;
if covarm[numvar,numvar]-som3<0 then begin
    listbox1.Items.add(timetostr(time)+' -->A matriz não é
positiva definida...');
    exit;
end else cholem[numvar,numvar]:=sqrt(covarm[numvar,numvar]-
som3);
for i:=1 to numvar do begin
    for j:=2 to numvar do begin

```

```

        if j>i then cholem[i,j]:=0;
    end;
end;
tarefas.Checked[5]:=true;
{calculando determinantes vivos}
listbox1.Items.Add(timetostr(time)+'    Calculando    determinantes
para vivos');
listbox1.Repaint;
deter:=1;
for k:=1 to numvar do begin
    deter:=deter*cholev[k,k];
end;
deter:=deter*deter;
tarefas.Checked[6]:=true;
listbox1.Items.Add(timetostr(time)+'    -->Determinante    para
vivos: '+floattostr(deter));
listbox1.Repaint;
////////////////////////////////////
{calculando determinantes mortos}
deter:=1;
for k:=1 to numvar do begin
    deter:=deter*cholem[k,k];
end;
deter:=deter*deter;
tarefas.Checked[6]:=true;
listbox1.Items.Add(timetostr(time)+'    -->Determinante    para
mortos: '+floattostr(deter));
listbox1.Repaint;
listbox1.Items.Add(timetostr(time)+'    Calculando    vetor    ao
quadrado para vivos');
listbox1.Repaint;
{calculando vetor ao quadrado vivos}
somav:=0;
for j:=1 to numvar do begin    ///multiplicação das matrizes
    res1[j]:=0;
    for k:=1 to numvar do begin
        res1[j]:=res1[j]+vetmdv[k]*inverv[k,j+numvar];
    end;
    somav:=somav+res1[j]*vetmdv[j];
end;
{calculando vetor ao quadrado mortos}
listbox1.Items.Add(timetostr(time)+'    Calculando    vetor    ao
quadrado para mortos');
listbox1.items.add(timetostr(time)+'    -->Vetor    médio    quadrado
vivos: '+floattostr(somav));
listbox1.Repaint;
somam:=0;
for j:=1 to numvar do begin    ///multiplicação das matrizes
    res1[j]:=0;
    for k:=1 to numvar do begin
        res1[j]:=res1[j]+vetmdm[k]*inverv[k,j+numvar];
    end;
    somam:=somam+res1[j]*vetmdm[j];
end;
listbox1.items.add(timetostr(time)+'    -->Vetor    médio    quadrado

```

```

mortos: '+floattostr(somam));
      {calculando coeficiente K}
      listbox1.Items.Add(timetostr(time)+'    Calculando    coeficiente
K');
      listbox1.Repaint;
      coefk:=0.5*ln(determ/determ)+0.5*(somav-somam);
      listbox1.Items.Add(timetostr(time)+'    -->Coeficiente    K:
'+floattostr(coefk));
      listbox1.Repaint;
      {calculando subtração das inversas das covariancias}
      listbox1.Items.Add(timetostr(time)+'    Calculando subtração das
inversas das covariâncias');
      listbox1.Repaint;
      for i:=1 to numvar do begin
        for j:=1 to numvar do begin
          difinv[i,j]:=inverv[i,j+numvar]-inverm[i,j+numvar];
        end;
      end;
      tarefas.checked[7]:=true;
      button2.Click;
    finally
      dbgrid1.DataSource:=datasource1;
      tabpes.filtered:=false;
    end;
end;

procedure Tfrmprj.TabPesAfterOpen(DataSet: TDataSet);
begin
  for numvar:=0 to 16 do if tabpes.Fields[numvar+2].value=null then
break;
end;

procedure Tfrmprj.ApplicationEvents1Exception(Sender: TObject;
E: Exception);
begin
  listbox1.Items.add(timetostr(time)+'    - '+e.Message);
end;

procedure Tfrmprj.Button4Click(Sender: TObject);
begin
  if messagedlg('Tem certeza que deseja apagar todos os
dados?',mtconfirmation,[mbyes,mbno],0)=mryes then begin
    que.SQL.Text:='delete          from          ORIGEM          where
grupes='+copy(combobox1.Text,1,1);
    que.ExecSQL;
    que.SQL.Clear;
    que.Close;
  end;
  tabpes.CommitUpdates;
  tabpes.Refresh;
end;

procedure Tfrmprj.Button2Click(Sender: TObject);
begin
  try

```

```

dbgrid1.DataSource:=datasource2;
priorim:=strtofloat(priori.text);
prioriv:=1-priorim;
listbox1.items.add(timetostr(time)+' Obtendo resultados para
vivos');
listbox1.items.add(timetostr(time)+' -->Ln(priorim/prioriv):
'+floattostr(ln(priorim/prioriv)));
listbox1.Repaint;
contvn:=0;
contvp:=0;
tabpes.filter:='grupes=''3'''; //validação vivos
tabpes.filtered:=true;
tabpes.First;
while not tabpes.eof do begin
    soml:=0;
    for j:=1 to numvar do begin          ///multiplicação das
matrizes
        res1[j]:=0;
        res2[j]:=0;
        res3[j]:=0;
        for k:=1 to numvar do begin

res1[j]:=res1[j]+(tabpes.Fields[k+1].AsFloat*difinv[k,j]);
            res2[j]:=res2[j]+vetmdv[k]*inverv[k,j+numvar];
            res3[j]:=res3[j]+vetmdm[k]*inverv[k,j+numvar];
        end;
        soml:=soml+(-0.5*res1[j]+res2[j]-
res3[j])*tabpes.fields[j+1].asfloat;
        end;
        soml:=soml-coefk-ln(priorim/prioriv);
        if soml>0 then inc(contvp) else inc(contvn);
        tabpes.next;
    end;
    resultado.Rows[1].Strings[1]:=inttostr(contvp);
    resultado.Rows[1].Strings[2]:=inttostr(contvn);
    listbox1.items.add(timetostr(time)+' Obtendo resultados para
mortos');
    listbox1.Repaint;
    contmn:=0;
    contmp:=0;
    tabpes.filter:='grupes=''4'''; //validação mortos
    tabpes.filtered:=true;
    tabpes.First;
    while not tabpes.eof do begin
        soml:=0;
        for j:=1 to numvar do begin          ///multiplicação das
matrizes
            res1[j]:=0;
            res2[j]:=0;
            res3[j]:=0;
            for k:=1 to numvar do begin

res1[j]:=res1[j]+(tabpes.Fields[k+1].asfloat*difinv[k,j]);
                res2[j]:=res2[j]+vetmdv[k]*inverv[k,j+numvar];
                res3[j]:=res3[j]+vetmdm[k]*inverv[k,j+numvar];
            end;
        end;
    end;
end;

```

```

        end;
        som1:=som1+(-0.5*res1[j]+res2[j]-
res3[j])*tabpes.fields[j+1].asfloat;
        end;
        som1:=som1-coefk-ln(priorim/prioriv);
        if som1>0 then inc(contmp) else inc(contmn);
        tabpes.next;
    end;
    resultado.Rows[2].Strings[1]:=inttostr(contmp);
    resultado.Rows[2].Strings[2]:=inttostr(contmn);

tx.Text:=floattostr((strtoint(resultado.Rows[1].Strings[2])+strtoint
(resultado.Rows[2].Strings[1]))*100/(strtoint(resultado.Rows[1].Stri
ngs[2])+strtoint(resultado.Rows[2].Strings[1])+strtoint(resultado.Ro
ws[2].Strings[2])+strtoint(resultado.Rows[1].Strings[1])));
tx.text:=copy(tx.text,1,pos(',',tx.text)+2)+'%';
listbox1.items.add(timetostr(time)+' Final da classificação');
listbox1.Repaint;
tabpes.filtered:=false;
finally
    dbgrid1.DataSource:=datasource1;
    tabpes.filtered:=false;
end;
end;

procedure Tfrmprpri.Button5Click(Sender: TObject);
begin
    priorim:=strtofloat(priori.text);
    prioriv:=1-priorim;
    listbox1.Items.add(timetostr(time)+'      Obtendo      resultado      da
classificação do indivíduo');
    som1:=0;
    for j:=1 to numvar do begin          ///multiplicação das matrizes
        res1[j]:=0;
        res2[j]:=0;
        res3[j]:=0;
        for k:=1 to numvar do begin

res1[j]:=res1[j]+(strtofloat(novo.Rows[1].strings[k])*difinv[k,j]);
            res2[j]:=res2[j]+vetmdv[k]*inerv[k,j+numvar];
            res3[j]:=res3[j]+vetmdm[k]*inverm[k,j+numvar];
        end;
        som1:=som1+(-0.5*res1[j]+res2[j]-
res3[j])*strtofloat(novo.Rows[1].strings[j]);
    end;
    som1:=som1-coefk-ln(priorim/prioriv);
    if som1>0 then begin
        resind.Text:='Indivíduo      classificado      na      população      com
probabilidade de sobrevivência';
        resind.Font.Color:=clblue;
    end else begin
        resind.text:='Indivíduo      classificado      na      população      com
probabilidade de óbito';
        resind.Font.color:=clred;
    end;
end;

```

```

end;

procedure Tfrmprpri.Button6Click(Sender: TObject);
var lis:tstringlist;
begin
  lis:=tstringlist.create;
  try
    lis.Add('priori : '+priori.text);
    lis.add('coefk : '+floattostr(coefk));
    lis.Add('vetmdv
'+floattostr(vetmdv[1])+';'+floattostr(vetmdv[2])+';'+floattostr(vet
mdv[3])+';'+floattostr(vetmdv[4])+';'+floattostr(vetmdv[5])+';'+floo
ttostr(vetmdv[6])+';'+floattostr(vetmdv[7])+';'+floattostr(vetmdv[8]
)+';'+floattostr(vetmdv[9])+';'+floattostr(vetmdv[10])+';'+floattost
r(vetmdv[11])+';'+floattostr(vetmdv[12])+';'+floattostr(vetmdv[13])+
';'+floattostr(vetmdv[14])+';'+floattostr(vetmdv[15])+';'+floattostr
(vetmdv[16])+';');
    lis.Add('vetmdm
'+floattostr(vetmdm[1])+';'+floattostr(vetmdm[2])+';'+floattostr(vet
mdm[3])+';'+floattostr(vetmdm[4])+';'+floattostr(vetmdm[5])+';'+floo
ttostr(vetmdm[6])+';'+floattostr(vetmdm[7])+';'+floattostr(vetmdm[8]
)+';'+floattostr(vetmdm[9])+';'+floattostr(vetmdm[10])+';'+floattost
r(vetmdm[11])+';'+floattostr(vetmdm[12])+';'+floattostr(vetmdm[13])+
';'+floattostr(vetmdm[14])+';'+floattostr(vetmdm[15])+';'+floattostr
(vetmdm[16])+';');
    for i:=1 to numvar do begin
      lis.add('difinv
'+floattostr(difinv[i,1])+';'+floattostr(difinv[i,2])+';'+floattostr
(difinv[i,3])+';'+floattostr(difinv[i,4])+';'+floattostr(difinv[i,5]
)+';'+floattostr(difinv[i,6])+';'+floattostr(difinv[i,7])+';'+floatt
ostr(difinv[i,8])+';'+floattostr(difinv[i,9])+';'+floattostr(difinv[
i,10])+';'+floattostr(difinv[i,11])+';'+floattostr(difinv[i,12])+';'
+floattostr(difinv[i,13])+';'+floattostr(difinv[i,14])+';'+floattost
r(difinv[i,15])+';'+floattostr(difinv[i,16])+';');
    end;
    for i:=1 to numvar do begin
      lis.add('inerv
'+floattostr(inerv[i,1+numvar])+';'+floattostr(inerv[i,2+numvar])+
';'+floattostr(inerv[i,3+numvar])+';'+floattostr(inerv[i,4+numvar]
)+';'+floattostr(inerv[i,5+numvar])+';'+floattostr(inerv[i,6+numva
r])+';'+floattostr(inerv[i,7+numvar])+';'+floattostr(inerv[i,8+num
var])+';'+floattostr(inerv[i,9+numvar])+';'+floattostr(inerv[i,10+
numvar])+';'+floattostr(inerv[i,11+numvar])+';'+floattostr(inerv[i
,12+numvar])+';'+floattostr(inerv[i,13+numvar])+';'+floattostr(inve
rv[i,14+numvar])+';'+floattostr(inerv[i,15+numvar])+';'+floattostr(
inerv[i,16+numvar])+';');
    end;
    for i:=1 to numvar do begin
      lis.add('inerm
'+floattostr(inerm[i,1+numvar])+';'+floattostr(inerm[i,2+numvar])+
';'+floattostr(inerm[i,3+numvar])+';'+floattostr(inerm[i,4+numvar]
)+';'+floattostr(inerm[i,5+numvar])+';'+floattostr(inerm[i,6+numva
r])+';'+floattostr(inerm[i,7+numvar])+';'+floattostr(inerm[i,8+num
var])+';'+floattostr(inerm[i,9+numvar])+';'+floattostr(inerm[i,10+
numvar])+';'+floattostr(inerm[i,11+numvar])+';'+floattostr(inerm[i

```

```

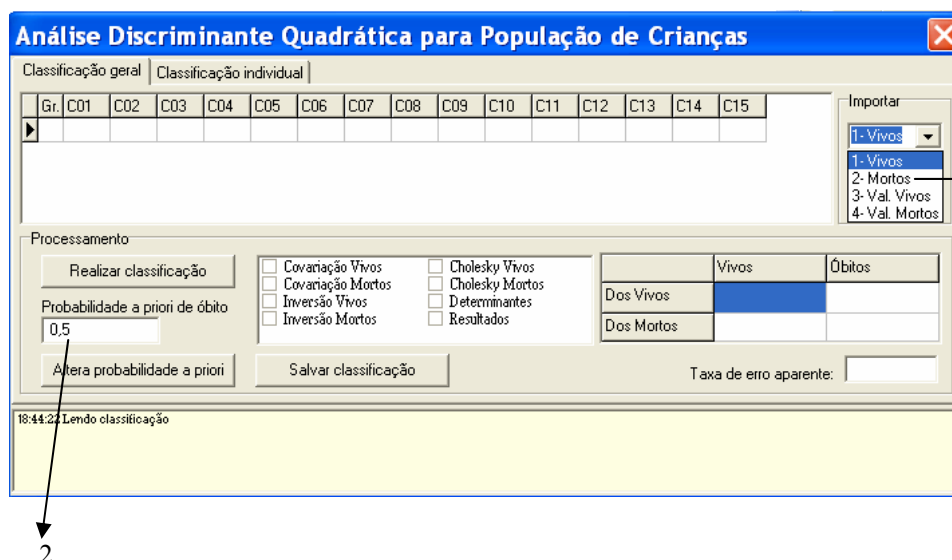
,12+numvar]))+';'+floattostr(inverm[i,13+numvar]))+';'+floattostr(inve
rm[i,14+numvar]))+';'+floattostr(inverm[i,15+numvar]))+';'+floattostr(
inverm[i,16+numvar]))+';');
    end;
    lis.SaveToFile('vars.ini');
    listBox1.Items.Add(timetostr(time)+' Classificação salva com
sucesso...')
    finally
        lis.free;
    end;
end;

end.

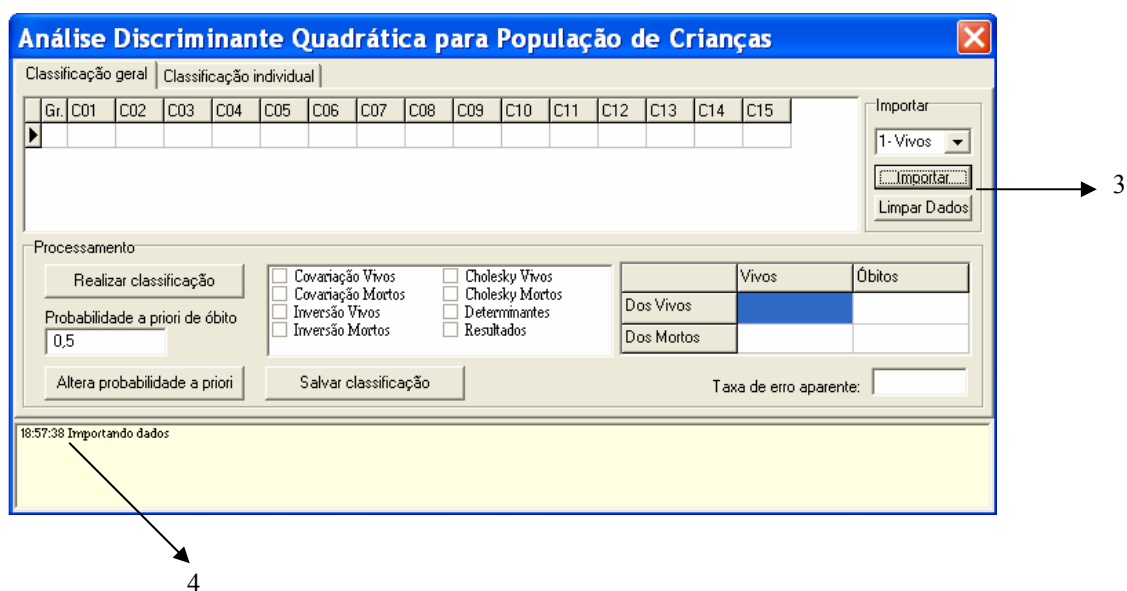
```


ANEXO IV – MANUAL DO PROGRAMA

Programa Desenvolvido – Tela 1: Classificação Geral

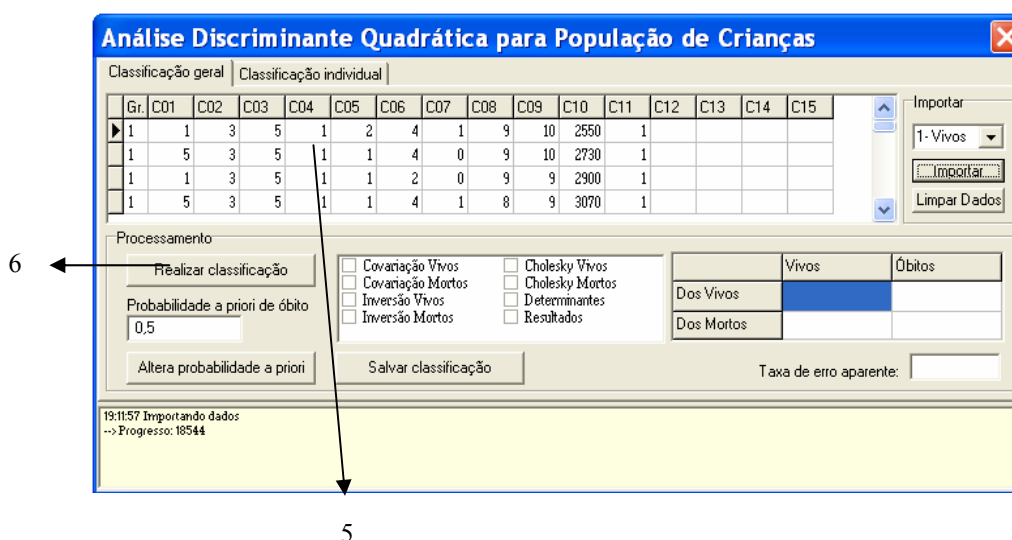


- 01) Escolha qual **matriz de dados** deseja importar ou excluir:
- 1- Matriz de Origem dos Dados: Crianças Sobreviventes;
 - 2- Matriz de Origem dos Dados: Crianças Não-Sobreviventes;
 - 3- Matriz de Validação da Função de Reconhecimento e Classificação: Crianças Sobreviventes;
 - 4- Matriz de Validação da Função de Reconhecimento e Classificação: Crianças Não-Sobreviventes.
- 02) Insira o valor da **probabilidade de óbito a priori**.



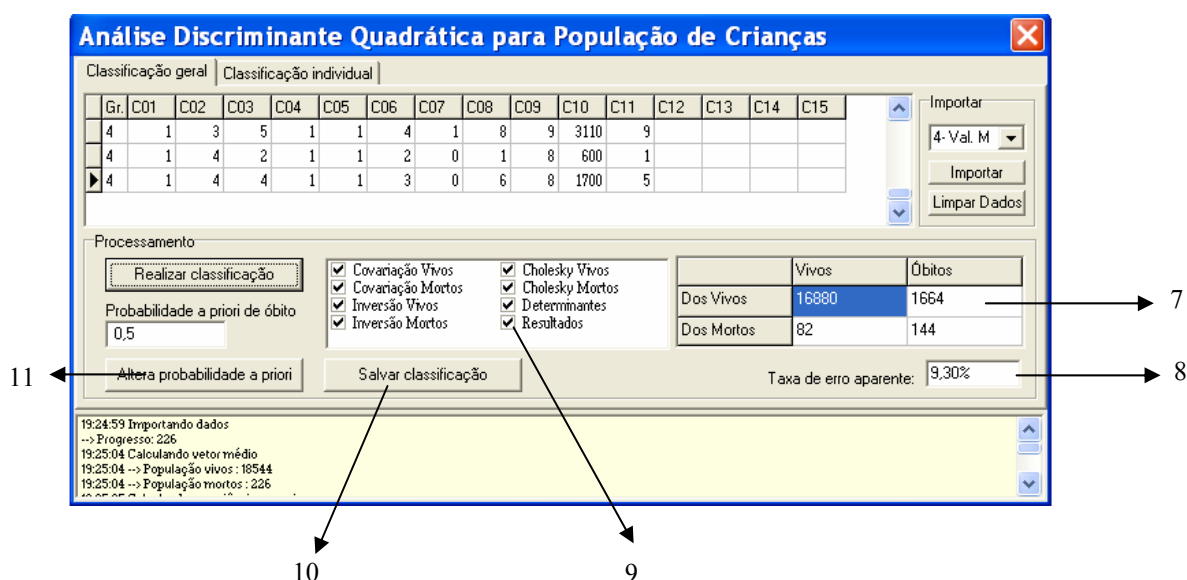
03) Escolha **importação** ou **exclusão** dos dados. No caso de importação, abre-se uma janela do Windows e então se escolhe o arquivo que contém a matriz de dados a ser utilizada. No caso de exclusão, surge uma mensagem de confirmação da exclusão dos dados. Os dados importados ou excluídos serão aqueles selecionados no item 01.

04) O **progresso da importação de dados** é apresentado neste campo. Neste campo também são mostradas outras informações sobre o processamento dos dados.



05) Os **dados importados** são mostrados nestas células.

06) Após importar todos os dados necessários para o funcionamento do programa, clique nesta tecla para iniciar o processo de construção da função de reconhecimento e classificação e de validação da mesma.



- 07) O **resultado da classificação** é apresentado neste campo.
- 08) A **avaliação da função de reconhecimento e classificação** (taxa de erro – aparente ou estimada) é indicada neste campo.
- 09) Os **processos completados** são indicados nesta área.
- 10) Para **salvar** a função de reconhecimento e classificação obtida e usá-la para classificar novos indivíduos, clique nesta tecla.
- 11) Para **alterar a probabilidade a priori** e realizar **nova classificação** dos dados basta trocar o valor do campo 02 e em seguida clicar nesta tecla

Programa Desenvolvido – Tela 2: Classificação Individual

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Dados da criança a ser classificada

Variáveis	Idade	Est.Civil	Escolaridade	Gestação	Gravidez	Parto	Consultas	Sexo	Apgar1	Apgar5	Pi
Valores											

Realizar classificação

Resultado:

19:24:59 Importando dados
 --> Progressor: 226
 19:25:04 Calculando vetor médio
 19:25:04 --> População vivos : 18544
 19:25:04 --> População mortos : 226

- 12) **Inserir os dados** referentes do novo indivíduo (criança) a ser classificada.
- 13) Clicar na tecla **realizar classificação** para proceder a classificação do novo indivíduo.

Análise Discriminante Quadrática para População de Crianças

Classificação geral | Classificação individual

Dados da criança a ser classificada

Variáveis	Escolaridade	Gestação	Gravidez	Parto	Consultas	Sexo	Ápgar1	Ápgar5	Peso	Nº Gestações
Valores	1	4	1	1	4	0	8	9	2500	4

< >

Realizar classificação

Resultado: Indivíduo classificado na população com probabilidade de óbito

19:24:59 Importando dados
 --> Progresso: 226
 19:25:04 Calculando vetor médio
 19:25:04 --> População vivos : 18544
 19:25:04 --> População mortos : 226

14

14) O **resultado da classificação** é apresentado neste campo. Para destacar, usou-se a cor vermelha para a classificação em óbito e a cor azul para a classificação em sobrevivência.